REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	31 August 2005	Final Technical Report (05NOV04 - 31AUG05)

4. TITLE AND SUBTITLE
System Dynamics Aviation Readiness Modeling Demonstration - Final Report

5. FUNDING NUMBERS
C N00014-04-F-0265

6. AUTHOR(S)
Michael E. McDevitt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CACI Dynamic Systems Inc., 10085 Scripps Ranch Ct., San Diego, CA 92131

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A proof-of-concept demonstration of System Dynamics modeling was developed to determine the relative merit of the approach and then an experiment compared the results of the System Dynamics model with a traditional linear regression readiness model. This report documents the proof of concept model development and the experiment's results. An aviation readiness production model was formulated using system dynamics. The model was developed over a period of six months with participation from subject matter experts at Commander Naval Air Force and Commander Strike Fighter Wing Pacific. The model incorporates monthly flight hour execution taking into consideration aircraft mission capable rates and the location of a squadron relative to the Fleet Response Plan (FRP) training cycle to generate Primary Mission Area points in strike warfare that are accumulated to generate an Attack Index. Concurrently, a renewed Strike Warfare Proficiency (Strike PRO) analysis was conducted for comparison. Twenty-four Carrier Air Wing (CVW) events at NSAWC Fallon, NV were used to benchmark performance. The results of the Strike PRO algorithm and the results of the System Dynamics Model were each compared against the performance metric. Source code for the System Dynamics Model is included as an appendix to the report.

14. SUBJECT TERMS
System Dynamics, Modeling, Linear Regression, Comparison, Aviation Readiness, Strike Warfare

15. NUMBER OF PAGES
63

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

CACI
EVER VIGILANT

System Dynamics Readiness Modeling Demonstration

Final Report

Developed for:

Mr. Ervin Kapos, Code N00F
Director, Operations Analysis Programs
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5660

August 31 2005

Work performed under Contract Number: N00014-04-F-0465

by

CACI Dynamic Systems Inc.
10085 Scripps Ranch Court
San Diego, CA 92131

UNCLASSIFIED

1.0 Introduction

CACI was contracted by the Office of Naval Research (ONR) to experiment with a non-traditional approach to improve readiness modeling in the Navy. CACI performed a proof-of-concept demonstration of System Dynamics modeling to determine the relative merit of the approach and then conducted an experiment to compare the results of the System Dynamics model with a traditional linear regression readiness model. This report documents the proof of concept model development and the experimental results.

2.0 Executive Summary

An aviation readiness production model was formulated using system dynamics. The model was developed over a period of six months with participation from subject matter experts at Commander Naval Air Force and Commander Strike Fighter Wing Pacific.

The model incorporates monthly flight hour execution taking into consideration aircraft mission capable rates and the location of a squadron relative to the Fleet Response Plan (FRP) training cycle to generate Primary Mission Area points in strike warfare that are accumulated to generate an Attack Index.

The development of the proof of concept systems dynamics model went well and the demonstration model received positive endorsement from the customer at Commander Naval Air Force for its potential utility in what-if analysis in regard to changing training and readiness requirements.

Concurrently, a renewed Strike Warfare Proficiency (Strike PRO) analysis was conducted for comparison. The bulk of this work involved data collection over a six-year period of disparate data sources. Twenty-four Carrier Air Wing (CVW) events at NSAWC Fallon NV were used to benchmark performance. The results of the Strike PRO algorithm and the results of the System Dynamics Model were each compared against the performance metric. They achieved almost equivalent statistical results, with correlation coefficients of 0.19 in both cases.

Therefore, the experiment was inconclusive. The analysis team was unable to either prove or disprove the null hypothesis. This was primarily due to historical data problems: timeframe mismatch, data quality and the non-availability of data. Both models may lack significant variables in their formulations, but data quality and scope over the performance period made any further analysis exceedingly difficult. On a positive note, the system dynamics model was qualitatively judged as having merit for providing a rational systems-level explanation for resource allocation in readiness production. Further use of the system dynamics method is recommended for analysis of training and readiness matrix requirements.

It is recommended that the Naval Aviation Enterprise take a close look at the requirements i.e., performance measures, methodology and culture associated with archiving linked sets of performance and resource data to enable current and future proficiency analyses. If the Navy leadership is serious about cost wise readiness, then some significant cultural practices in the way that individual performance data are collected, archived and managed will need major revision and in some cases, initial implementation.

3.0 Background

This effort was initially proposed in an unsolicited white paper presented to Mr. Ervin Kapos, Director of Operations Research to the Chief of Naval Research in the Spring of 2003. The rationale for the demonstration is included in the following problem statement and proposed technical approach.

3.1 Problem Statement

Recent newspaper articles report that the Chief of Naval Operations is planning a major overhaul of maintenance, training and deployment programs to quicken the Navy's response. "A force that is ready every day to apply maximum combat power...with strike groups fully ready to go out on very little notice."[1] This radical redesign of how forces are trained, maintained and employed could mean that traditional concepts of readiness modeling may fail to provide the right information required by decision-makers. This should prompt the question, do we need to relook at how we model readiness.

With the advent of a new defense readiness reporting system (DRRS) requirement from OSD there is an urgent need for readiness capability models (by primary mission area) that can provide joint force commanders with an accurate assessment of a unit's capability to perform specific Joint Mission Essential Task List (JMETL) requirements now and in the future. This assessment methodology must have credible analytical underpinnings to allow for proper force selection and to calculate a realistic time to get ready to perform the mission and the time-associated costs. The new methodology must base decisions on currently available data and databases.

A "useful" readiness model should be able to provide the Joint Force Commander with an assessment of a unit's capability to perform mission essential tasks and that can answer the following questions:

- Ready to do what? At what level of performance, in what conditions, and for how long?
- Ready when? Today? Next Month? After mobilization?
- How much time and how much will it cost to achieve a desired level of readiness – to get ready?
- Ready for how long? What is the sustained capacity to perform the mission?

Current readiness assessment is based on the Status Of Readiness and Training (SORTS) paradigm, which provides a "perceived" condition of a unit as measured in four resource categories: Personnel (CRPER), Equipment (CREQP), Training (CRTNG) and Supply (CRSUP). These categories are then rolled up by mission area to provide a Primary Mission Area (PRMAR) rating of M-1 to M-4, with M-1 being "fully ready" and M-4 being "not ready". Each resource area is also rated in a similar fashion with "C-ratings", C-1 through C-4. All of the primary mission areas and resource categories are compared and the worst rating generally equates to an overall C-rating or CROVL for the unit.

Unfortunately with the current system, with the exception of narrative comments, the time and resources required for a unit to achieve the next level of readiness or the final level of readiness is usually not specified, making it harder for joint force planners to determine ultimate mission readiness at some later date. When readiness changes, a new report is submitted. Readiness reporting tends not to be continuous, but instead is event-based and quantized. Trends are not explicitly considered nor reported unless included by a thoughtful commander as a remark.

---

[1] "Navy plans to alter training, quicken response times", San Diego Union Tribune, Joe Cantlupe , Copley News Service, 6 April 2003

The current system is based on unit commanders' reports of current status according to certain formula designated by type commanders. Depending on the type commander and the culture of the unit type there may be considerable variability in the detail and methodology substantiating the report. There have also been cases of reporting a condition by default. A deployed unit must be C-2 to deploy so a unit reports C-2 just prior to deploying. For example, cancellation of casualty reports prior to deployment was routinely followed by new casualty reports on the same equipment a few days after the deployment begins.

Current readiness systems ignore the persistence of readiness and the implications of inertia when considering state change. Once a unit is ready, except for restocking of consumables, the unit will tend to remain ready for a sustained period. Readiness will slowly decline if the skills and practices in the mission are not refreshed at some periodicity. These erosion curves tend to hold true if the personnel dimension and other factors remain relatively constant. Performance of the mission task increases readiness - up to the point where fatigue begins to affect performance and increase error rate.

Key to the overall "readiness system" is that there is considerable feedback present between the various resource elements that contribute to readiness. Personnel shortages impact training. Training can be impacted by equipment and supply. Equipment readiness can be impacted by training or a shortage of qualified personnel. There are variable delays in acquiring training. A delay in readiness improvement depends on the current state of readiness.

The key metrics in a readiness system should be "total time to get ready" and "total cost to get ready" for equivalent levels of mission capability or performance. Recent experience in preparing ships, air wings and battle groups for accelerated deployment in support of the "War in Iraq" may provide ample time-series data for examination of these concepts.

3.2 Technical Approach

Current readiness estimation techniques neglect the necessity of having a continuous flow of information from the vehicle level including the human performance portion of weapons' system(s) through to the theater level. Concepts like operational based readiness assessment (OBRA) attempted to measure various readiness indicators and develop a concept of continuous systems availability as a measure of readiness. However, efforts were compounded by a systemic lack of periodic measurement of performance information and the ability to correlate those indicators reliably to performance. Additionally, the coupling of resource information to performance is imprecise across many if not all of the resource domains.

Could it be that <u>the analysts are using the wrong tools?</u> Linear regression models (LRM) tend to become curve-fitting exercises with adjustment of constants. Over time as the LRM fit erodes, the constants are readjusted, but with little justification. Excel spreadsheets fail to account for feedback and time delays. Modern system dynamics software provides new tools and a refreshing new look at using cybernetic theory to explain readiness behaviors.

Readiness tends to behave in a cyclic fashion for a number of reasons. One reason is that the readiness "system" is designed that way. A graphic example of cyclic readiness, the readiness of non-deployed air wings over time, is shown in Figure 1.

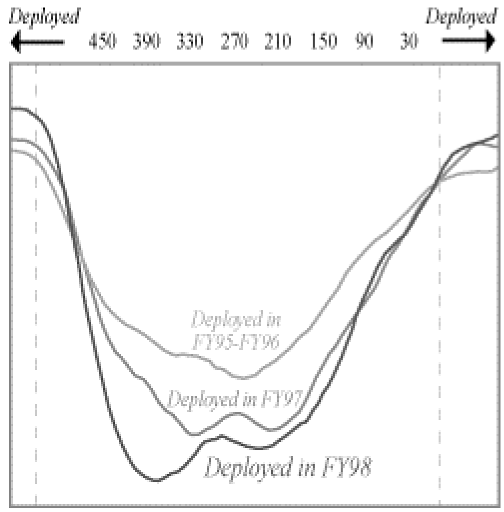


Figure 1: Non-deployed Air Wing Readiness "Bathtub Chart"

Following a deployment, units enter the Inter Deployment Training Cycle (IDTC) – now called the Fleet Readiness Training Plan (FRP). During early stages of the FRP, readiness degradation is expected as ships and aircraft undergo maintenance and crews turnover. As units progress through the FRP, readiness should steadily improve as maintenance is completed and training opportunities increase. In the latter stages of the FRP, units hone their war fighting skills by participating in exercises designed to ensure full combat readiness prior to deployment.

Non-deployed readiness is currently funded at levels that leave little margin for flexibility. When funding shortfalls occur, the Navy focuses first on ensuring the full readiness of deployed forces. Consequently, non-deployed readiness suffers as units in earlier stages of the FRP defer the ordering of parts, maintenance, and training so that additional funds can be made available for deployed units. While this allows us to maintain a satisfactory deployed readiness posture, it has an undesirable effect on non-deployed forces. As a result, the non-deployed Air Wing readiness curve, or "bathtub," normally associated with units in the FRP has become increasingly deep and recovery to full combat readiness has become much more taxing and occurs later in the FRP. The "bathtub" chart illustrates the increasing difficulty our non-deployed forces are experiencing as they pass through the FRP.

The results of a simple cyclic readiness model using a continuous time simulation methodology from System Dynamics recreate the “pattern of readiness” evidenced in the bathtub effect as shown in Figure 2.

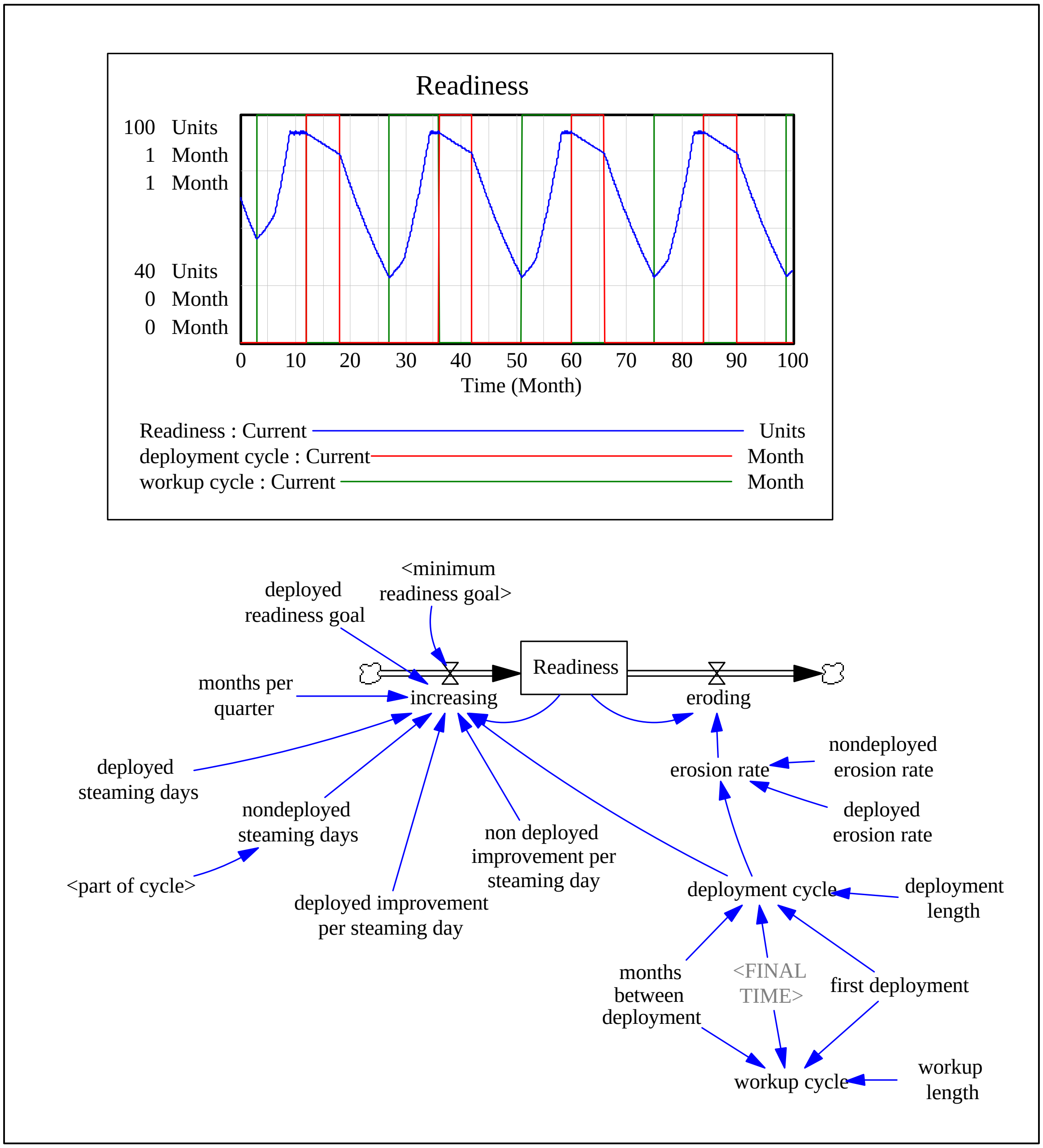


Figure 2: Cyclic Readiness Model

The depth of the bathtub and relative steepness of the slope to climb out increase and decrease based on policy decisions and the dynamic availability of resources. As fewer resources are available for training after return from deployment, the training levels drop lower until the resources are finally made available late in the workup cycle. This effect became very evident in the mid-1990s in Naval Aviation when airframes were unavailable for workup training due to modernization (P-3 ICAPS) or due to significant maintenance problems (F/A-18 gas turbine engine shortages).

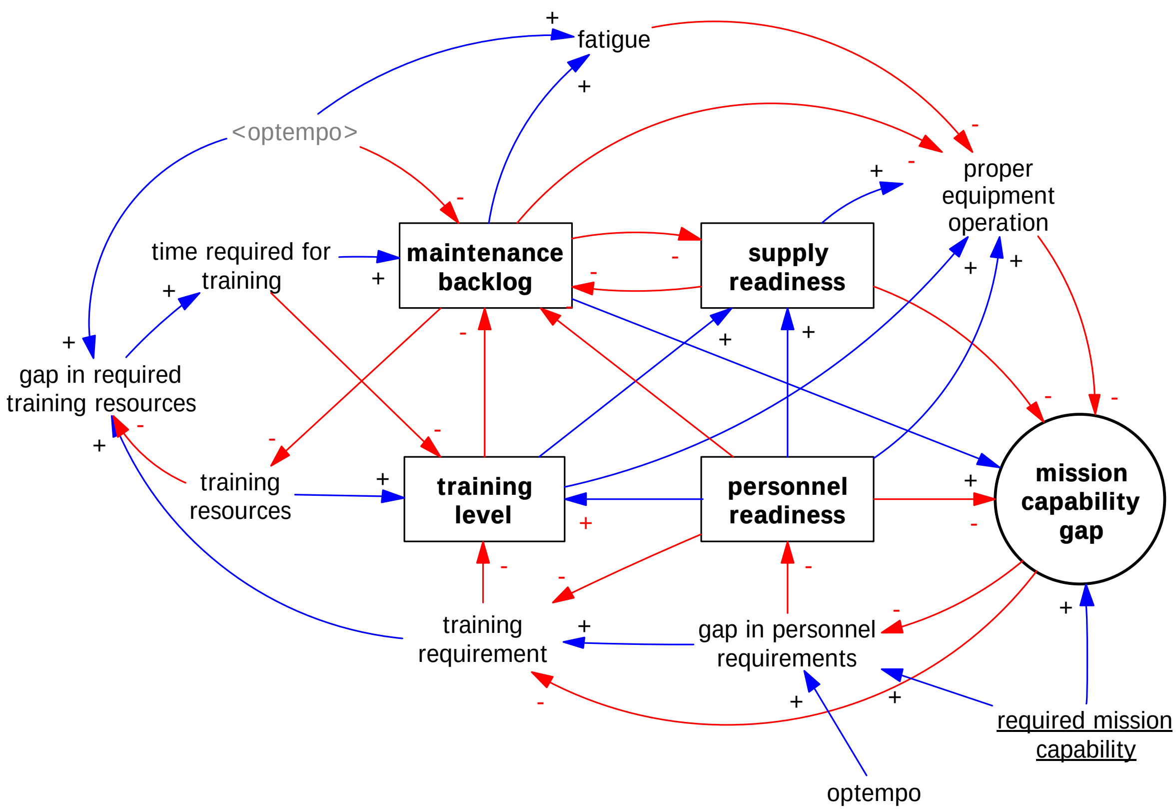


Figure 3: Causal Loop Diagram[2]

Figure 3 shows a causal loop diagram that illustrates the dynamic feedback between the traditional readiness resources of personnel (CRPERS), training (CRTNG), equipment (CREQP), and supply (CRSUP) as they impact readiness. Readiness in this context is measured as the level of performance of a warfare mission over a time period or as a clearly defined mission capability.

There are multiple non-linear relationships between and among the four major categories of readiness that are magnified by external influences, such as changes in operational tempo or budget, as they impact mission capability.

<u>Experimental Questions</u>: 1. Can System Dynamics modeling be used to address "Human in the Loop" problems implicit in current readiness assessment models? 2. Can a continuous time model with causal feedback loops and delays perform any better in readiness assessment than a linear regression model?

This demonstration/experiment considered readiness as an "inventory" that accrues over time and that can be reduced through expenditure and then restocked through the actions of a complex network of supply

---

[2] **Reading Causal Loop Diagrams (CLD).** Arrows ? that have no "sign" are assumed to be +. Negative relationships are marked with a – sign. + Relationships should be interpreted as follows: A? +B means as A increases, then all else being equal, B increases. OR, as A decreases, then all else being equal B decreases. They move in the SAME direction. – Relationships are interpreted as follows: A? -B means as A increases, then all else being equal, B decreases. OR, as A decreases, then all else being equal B increases. The two variables move in OPPOSITE directions.

chains. Readiness has persistence, or inertia when considered as an inventory. As an inventory, it can be modeled using supply chain theory and continuous time simulation techniques.

4.0 Statement of Work:

The contract authorized a proof of concept demonstration of the feasibility and potential benefits of using Systems Dynamics Readiness Modeling through CACI performance the following tasks:

Task 1: The Contractor shall conduct a proof of concept demonstration of the feasibility and potential of using Systems Dynamics Readiness Modeling to model readiness, as measured by operational performance.

Task 2: The Contractor shall develop a system dynamics readiness model for a selected primary mission area in consultation with the ONR Operations Analysis Program Office.

Task 3: The Contractor shall conduct an experiment to compare Systems Dynamics Readiness Modeling with linear regression-based readiness estimating models

Task 4: The Contractor will provide a formal written report (final report) that describes the results of the experiment as well as lessons learned and recommendations.

5.0 Research Project Plan

This project builds upon past efforts. The Strike Warfare Proficiency (Strike PRO) program was sponsored by CINCPACFLT N83 in the late 1990s. Strike PRO used a series of linear regression models to predict point performance of an Air Wing or F/A-18 squadron in the land attack mission.

5.1 Warfare Mission

Navy F/A-18 Attack squadrons in a Strike Warfare Mission were selected as the targeted area for the demonstration/experiment. Attack performance in Strike Warfare (STW) will be measured using the standard operational metric of Bomb Hits Assessed (BHA) divided by the number of Bombs Fragged (BF) or (BHA/BF)[3].

5.2 Methodology

Develop a system dynamics (SD) or continuous time simulation (CTS) model of naval aviation readiness based on "P4 + WARTS"[4] resource patterns as primary causal factors of mission readiness. Include and incorporate the "Thomas Group" metrics wherever appropriate. Specifically address Strike Warfare mission performance for the F/A-18 Type/Model/Series of aircraft.

1.1. Conduct a problem definition session at CNAP with appropriate process owners.

---

[3] StrikePRO used the number of bombs assessed as successfully hitting the target divided by the number of bombs that an air wing launched while on its Fallon detachment – BHA/BL. Later, on the recommendation of NSAWC SMEs, it was changed to BHA/BF – or the number of bomb hits divided by the number of bombs the air wing planned to drop (or bombs fragged)

[4] Originally coined at COMNAVAIRPAC in the mid-90s. P4+WARTS stands for the resources required to build and support aviation readiness from the Type Commander's perspective. These resources include Personnel, Planes, Parts, Petrol, Weapons, Adversaries, Ranges, TAD, and Simulators

1.1.1. Describe problem behavior verbally and graphically. Describe data sources and reliability. Describe goals, historic, current and projected. Explain rationale behind behavior curve.

1.1.2. Assess level of understanding and degree of confidence that the causes of the behavior are understood. Document causal relationships. Determine or establish consensus on causal factors and effects.

1.1.3. Define all pertinent metrics. Determine which metrics that are currently collected are control metrics. Indicate data sources and how frequently data are refreshed.

2. Collect data from identified data sources for model validation and verification.
3. Perform Model validation with aviation process owners.
4. Compare SD Model results with Strike PRO regression analysis model. Null Hypothesis: <u>That the system dynamics model will out perform the regression model by establishing a confidence of 0.01 or better.</u> The historical performance of the Strike PRO regression model is no better that .05.
5. Prepare and submit a formal written report that identifies the results of the experiment as well as lessons learned and recommendations.

5.3 Data sources

The following data sources will need to be accessed as <u>a minimum</u>:

- Aviation Data Warehouse (ADW) using the SHARP software program for data reporting.
- ACES – Navy Flight hour program budgets and execution data
- TRMS – Personnel and SORTS data
- Operational results from SFARP, NSWAC, C2EX, JTFEX and deployed.
- Past and present Training & Readiness instructions and matrices for the F/A-18

5.4 Metrics, units, time series

All metrics and data will need to be normalized across the FRP. For example, Bomb Hits Assessed (BHA) / Bombs Fragged (BF) across an FRP with the FRP cycles normalized to the same relative time scale, i.e., Month 0, 1, …18. The nomenclature used in the NAVRIIP program for example is based on D, where D+0 is the deployment month. D-6 is six months prior to the deployment and D+6 is six months into the deployment.

5.5 Detailed project plan

The actual project proceeded as shown in Figure 4.

ID	Task Name
1	**ONR Readiness Experiment**
2	Contract Award
3	Executive level brief at CNAP
4	Conduct a problem definition session at CNAP with appropriate process owners.
5	Describe problem behavior verbally and graphically.
8	Document causal relationships.
9	Receive DD254 from ONR
10	Describe data sources and reliability.
13	Collect data from identified data sources for model validation and verification.
14	Model Aviation Readiness
15	Perform Model validation with aviation process owners.
16	Compare SD Model results with Strike PRO regression analysis model.
17	Draft Report
18	Out Briefing Report Review of Findings
19	Finalize Report
20	Prepare and submit a formal written report that identifies the results of the experiment as well as

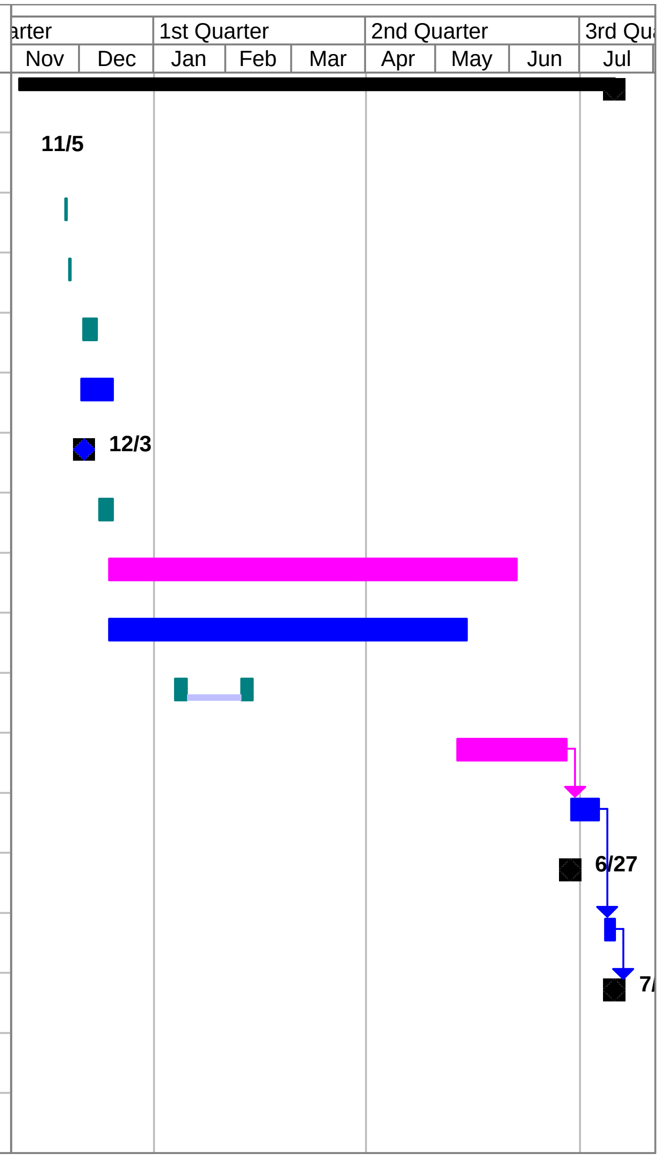


Figure 4: Project Plan

6.0 System Dynamics Aviation Readiness Production Model

The aviation readiness production model (or flight hour model) that was developed for this proof of concept demonstration is included in its entirety (complete with source code) in Appendix A. A detailed description of the model follows.

The model was developed on the basic premise that as hours were flown based on the tenets of the training and readiness (T&R) matrix an increase in readiness would accrue over time based upon the Navy's business rules and the general nature of human learning, i.e. specifically learning curves.

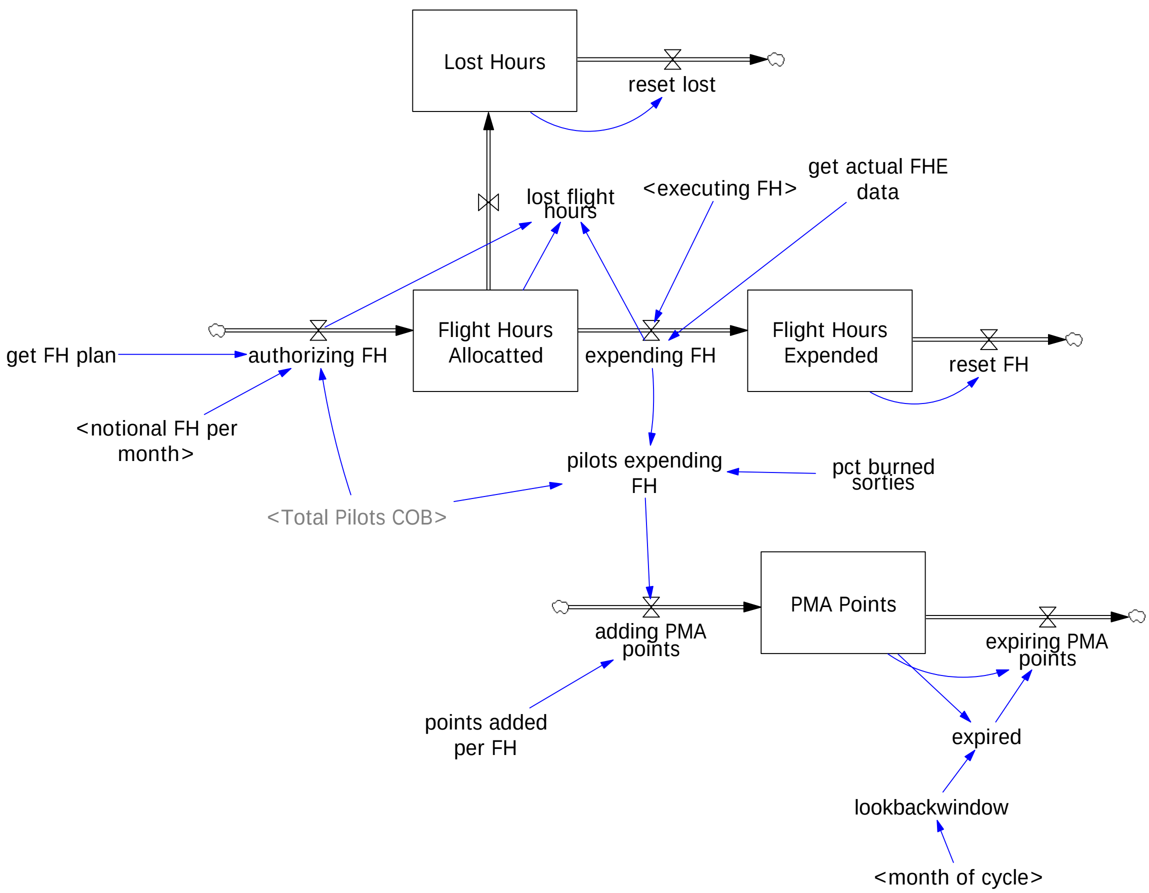


Figure 5: Flight Hour Expenditure

Figure 5 depicts the model structure for the planning and expenditure of flying hours and the accrual of Primary Mission Area (PMA) points. The Flight Hour Plan is determined based on the Notional Fleet Readiness Training Plan (FRTP) as shown in Figure 6 (Enclosure 4 of CNAF 3500.1B dated 21September 2004 - T&R Matrix). It is keyed to deployment cycles. Actual Air Wing/Squadron historical schedules, Flight Hour Execution and Full Mission Capable (FMC) rate data were fed to the model from an Excel spreadsheet.

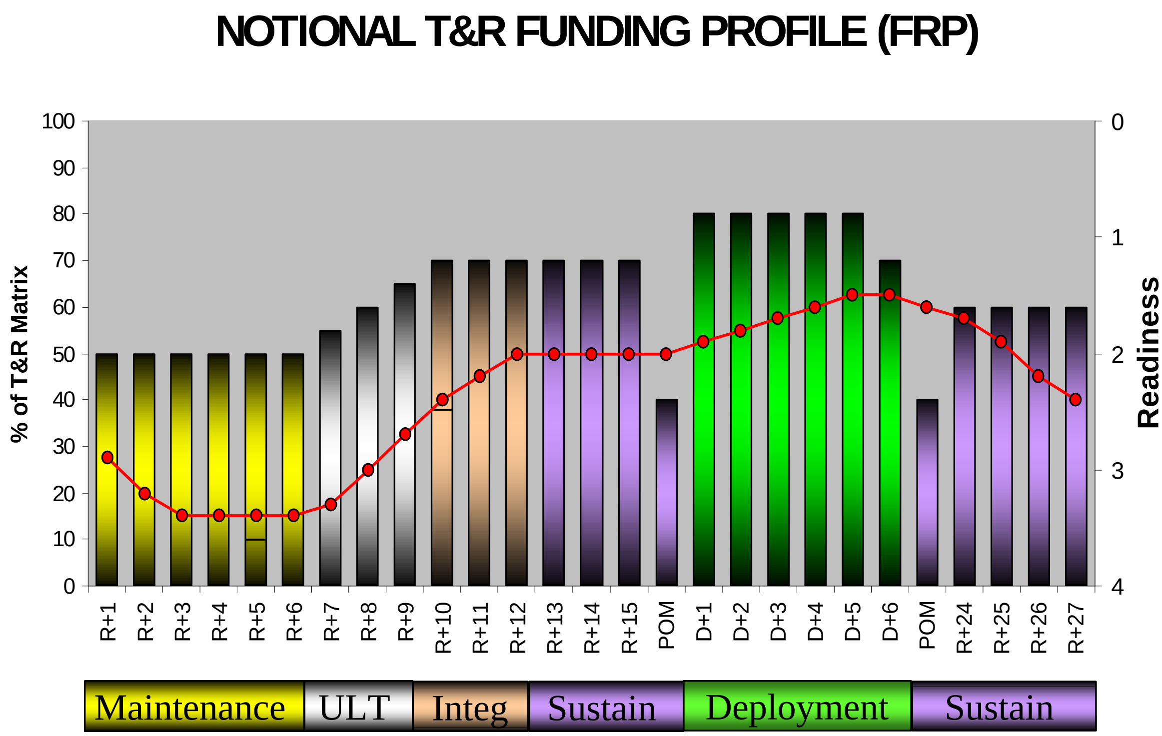


Figure 6: Notional T&R Funding Profile

The notional Flight Hours per month used in the model is based on the required number of Training and Readiness sorties multiplied by the number of flight hours (1.5) per F/A-18 sortie, all divided by the training interval of 90 days. This results in twenty-seven (27) flight hours per pilot and is used to calculate the Flight Hour Plan.

As actual flight hours are executed, a portion of these hours are allocated to training necessary for Strike Warfare. As Strike Warfare accounts for roughly a third of the hours required to be flown in accordance with CNAFINST (see Figure 8), thirty-three (33) points are added for each flight hour executed by full mission capable aircraft.

PMA points accrue and are maintained over the 27 Month FRTP cycle. At the beginning of each cycle the look back window is set to 90 days. This look back window expands up to 360 days, day per day after completion of the initial ninety-day training interval. As training events age beyond 360 days they expire and are required to be repeated. At the commencement of a new cycle, the look back window is again reset to ninety days.

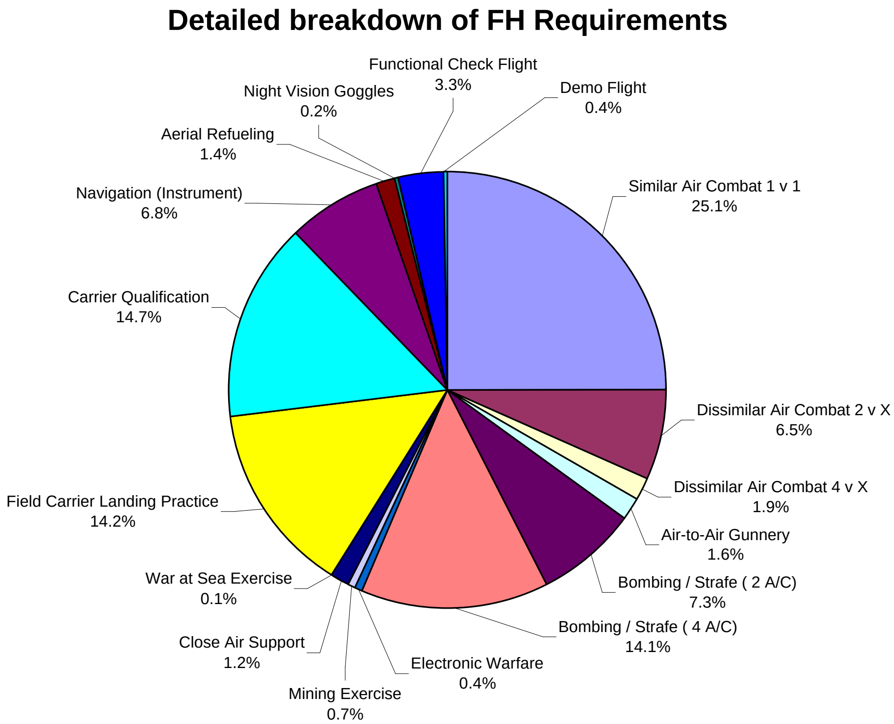


Figure 7: T&R Matrix Requirements for F/A-18

A simple supply chain monitors the number of pilots by level for each squadron in Figure 8. Replacement pilots are received from the Fleet Replacement Squadron (FRS) at about 3 every two months. They require on average six months to qualify as a Level 2 pilot. An additional year is required before qualification as a positional pilot. Another year is required to qualify as an advanced pilot capable of being in a strike lead role.

The model calculates the demand on the FRS based on gaps and expected losses. This model does not include the FRS pipeline or its associated pipeline delay. It also assumes that losses at the department head and command levels are filled immediately with contact relief and no delays in reporting.

The model was run assuming steady state pilot production that could easily support 15 pilots per squadron across the force. Unfortunately, historical squadron manning and monthly attrition data was not available for the time periods under analysis. The model is fully capable of taking time series attrition data into account.

The model continually calculates the number of pilots currently on board (COB) and keeps track of the various experience levels and the aggregate experience of the squadron. Percentage of L2, L3 and L4 pilots onboard is continuously monitored and can be readily compared against the defined readiness standards shown in Table 4:

ACTC	M1	M2	M3
>=L2	82%	75%	65%
>=L3	50%	45%	40%
>=L4	30%	25%	20%

Table 4: ACTC Experience Levels and M-Ratings

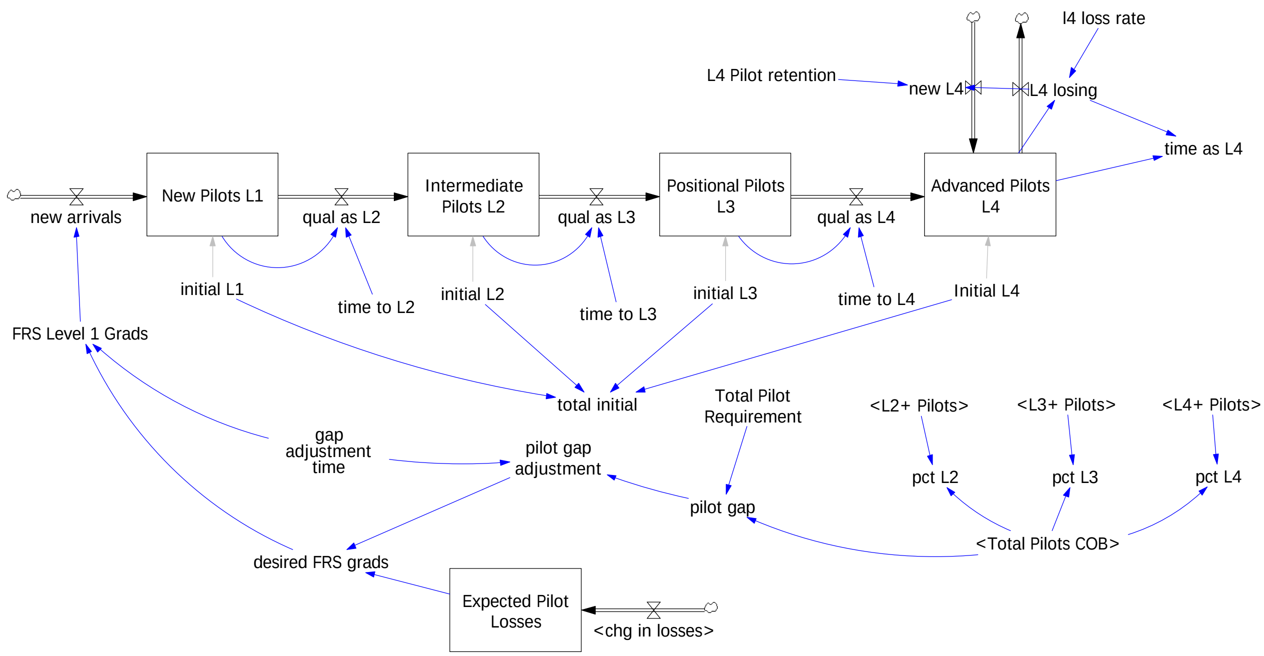


Figure 8: Pilot Supply Chain

The system dynamics model has the capability to use a look up feature to model non-linear relationships. In Figure 9, as the FRP cycle progresses from early to advanced the PMA points accumulated per flight hour that are added to the Attack Proficiency (AP) Units are reduced. In effect, the value of later training on proficiency is incrementally smaller later in the training cycle. Figure 10 shows the structure for adding to proficiency. Cycle time is calculated based on deployment schedules as shown in Figure 11.

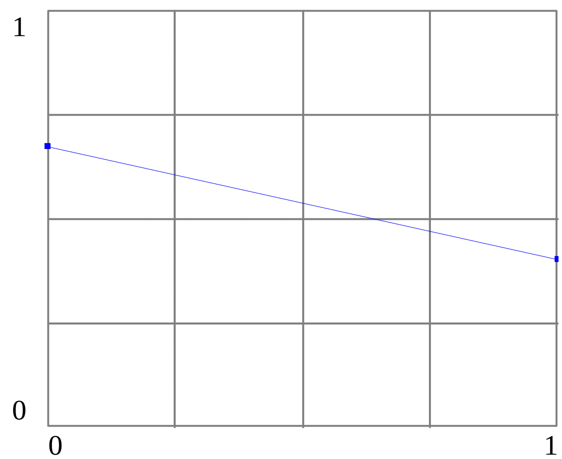


Figure 9: PMA Points by Cycle Position.

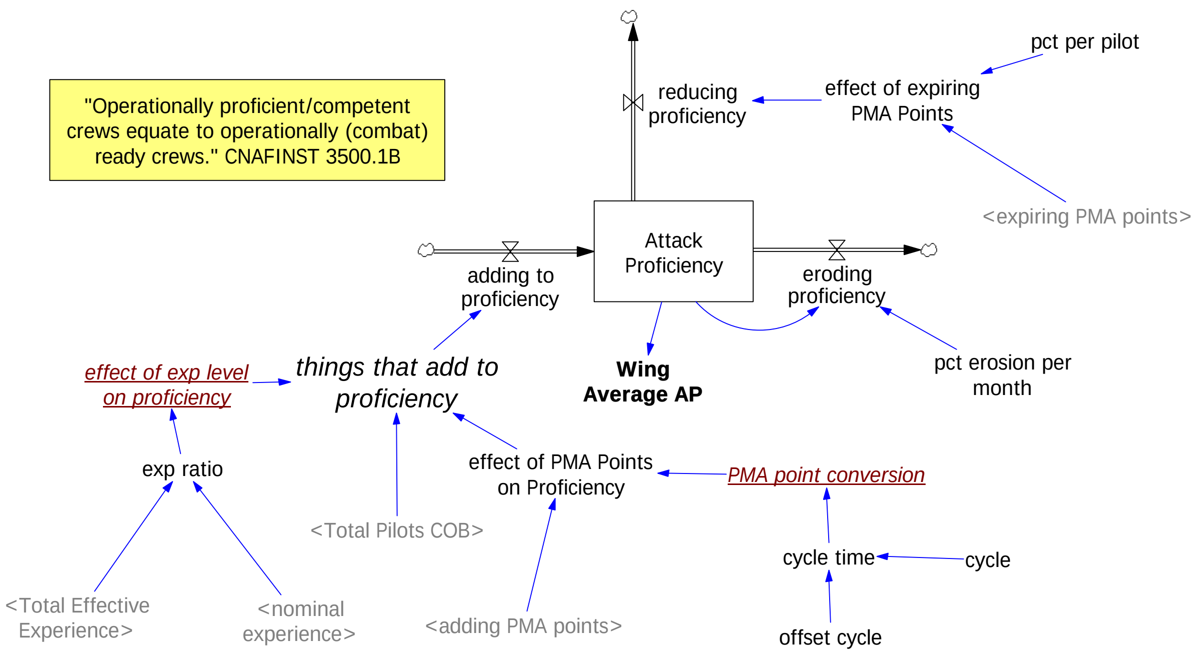


Figure 10: Attack Proficiency (AP)Index

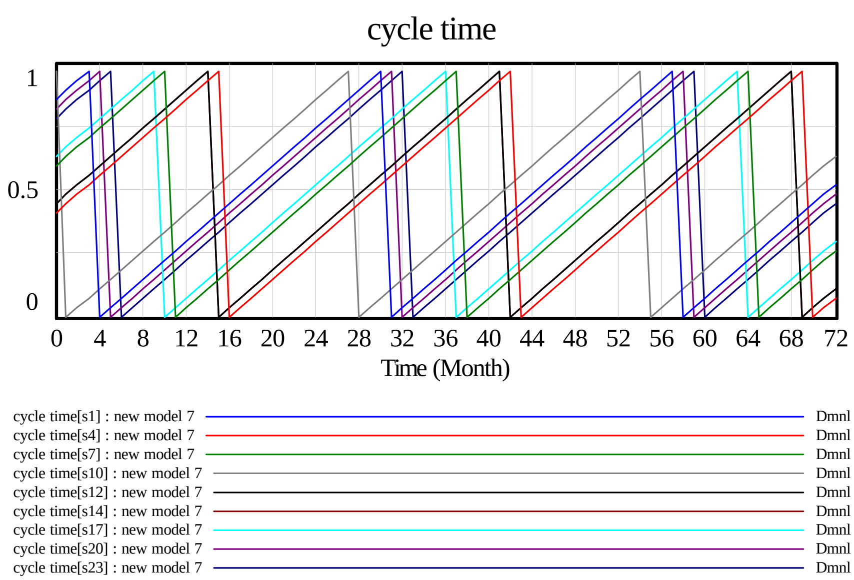


Figure 11: Cycle time by Wing

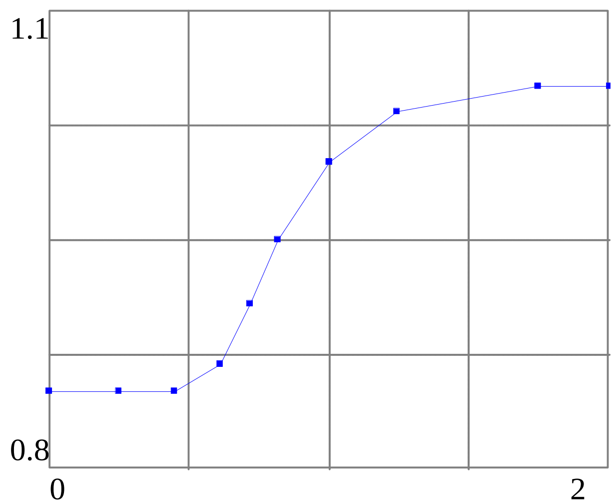


Figure 12: Experience level effect on gaining proficiency.

Figure 12 shows the relationship of squadron experience level on gaining proficiency, the more experienced a squadron is the more proficiency can be achieved per training event. This factor comes in to play when a squadron has large number of junior personnel.

7.0 Results

The experiment was inconclusive – we were unable to prove or disprove the null hypothesis. Results were that both models (System Dynamics and Linear Regression) had Pearson correlation coefficients of 0.19. This means that both models had confidence factors around 0.85. In other word, both models operated equally poorly and below expectations as detailed in Figure 13. Figure 13 compares the two models against the BHA/BF scores.

The theoretical model shows some promise – however – verification may prove costly and time consuming. As an example, note the polynomial "trend" of the BHA Scores depicted in Figure 13. Qualitatively, although the linear regression model had a much tighter fit, it was unable to accommodate the apparent sinusoidal pattern in the BHA scores as well as the System Dynamics Model. The seasonal oscillation trend in performance data can be readily explained by the causal, system dynamics model as shown in Figure 14. The ability of the SD model to replicate cyclical patterns of behavior over time supported CNAF and Strike Wing staff judgments that the system dynamics model had merit in that it provided a rational systems level explanation for resource allocation in readiness production. Further use of the system dynamics method is recommended for analysis of training and readiness matrix requirements.

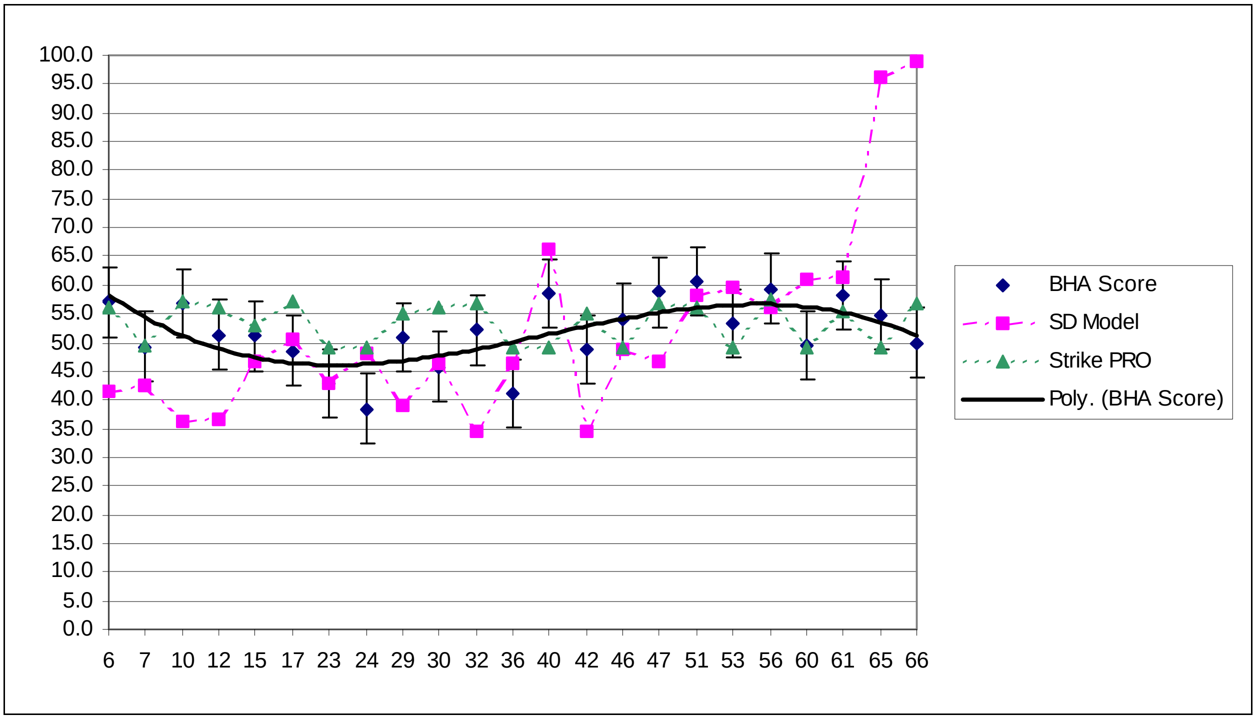


Figure 13: Comparison of Performance Models (Time in months)

Multiple R	0.19
R Square	0.04
Adjusted R Square	-0.11
Standard Error	0.07
Significance F	0.86

Table 3: Strike PRO Regression Statistics

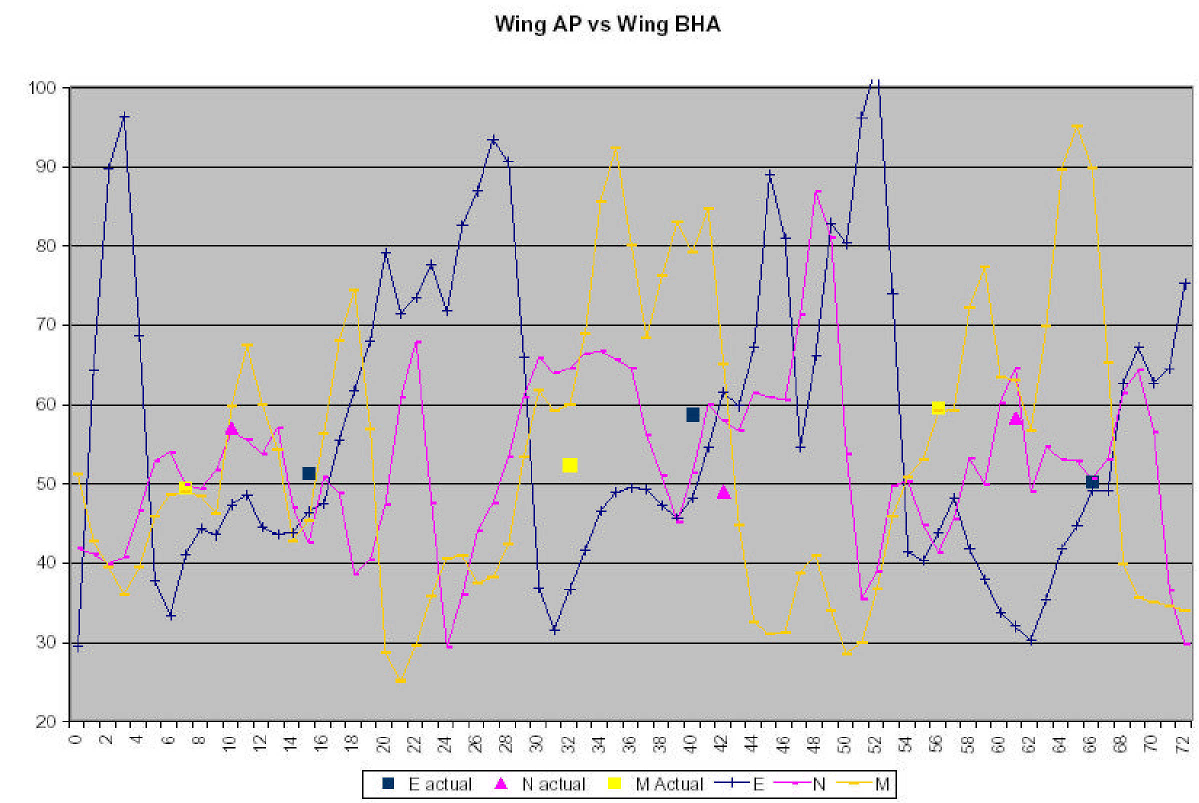


Figure 14: Wing Attack Proficiency (AP) plotted over time

8.0 Problems encountered

The primary problem plaguing the project was the inability to match historical performance data with resource data. The main reason for this data mismatch was mainly the time lag involved – performance data from the previous decade were unsupported by enough resource data from the same time period – a decade later. Attempts to get current performance data to match with current resource data was not supported by the Echelon Two Navy command that possessed the operational performance data. Additionally there were large gaps in the historical data as well as inherent errors in the data stream. In some cases the desired resource data was either not collected or not collected at the level of detail required to support the analysis.

The following paragraphs summarize some of the particular data related issues.

8.1 Data Problems

There were four sources for resource data. We experienced varying degrees of success in capturing data from each. Since the 24 air wing rotations at Fallon occurred between early 1996 and late 2001, we anticipated some difficulty in finding archived data that was up to nine years old.

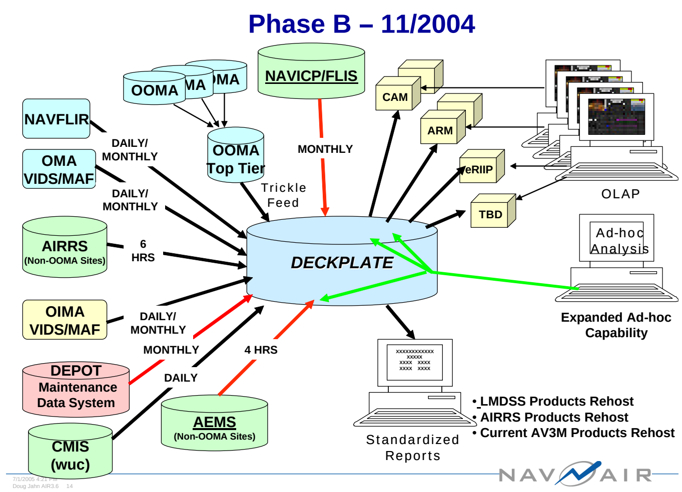
- Aviation Data Warehouse. Innova maintains ADW for CNAF on its SIPRNET site. While we were unable to mine the data we needed from their website, they were able to send us archived data that was kept offline that we used to explore F/A-18 squadron SORTS reports just prior to their Fallon

detachments. 19 of the requested 69 data points were missing, requiring us to make assumptions that call into question the validity of E2 data. We were unable to extract E3 data from this data.

- TRMS. Innova maintains TRMS for CNAF & CNAL on their SIPRNET websites. While we were unable to mine the data we needed from their website, they were able to send us archived data that was kept offline that we used to explore F/A-18 squadron personnel and manning status prior to their Fallon detachments. However, we were unable to reconstruct an adequately reliable or consistent picture of the personnel variables from the F/A-18 squadrons and consequently used none of the bricks within the P resource area.

- ACES. CACI provided Aviation Cost Evaluation System (ACES) flight hour data on the F/A-18 squadrons. The flight hour data from ACES differed from the flight hour data from ADW that differed from the flight hour data from the Naval Air Logistics Data Analysis (NALDA) program at NAVAIRSYSCOM. We estimate that there was an average error rate of 4% before we made assumptions that call into question the validity of K1 & M1 data.

- NALDA. By the end of the project, we were able to use the Decision Knowledge Programming for Logistics Analysis and Technical Evaluation (DECKPLATE) application to access the flight hour, Subsystem Capability Impact Reporting (SCIR), and maintenance databases -- Aircraft Inventory Readiness and Reporting System (AIRRS), Naval Aviation Maintenance and Material Management Yellow Sheet Reports (AV3M), Logistics Management Decision Support System (LMDSS), and Configuration Management Information System Work Unit Codes (CMIS WUC) at NALDA. Additionally, they were able to send CDR Ted Kaehler (CNAF N42) archived data that was intended to meet our needs. Unfortunately, the problem was not an inability to call up monthly data for specific Navy and Marine Corps F/A-18 squadrons from nine years ago. The problem was that about 4% of those months had no data reported or had data that was very obviously wrong. We used what judgment we could to massage the data, but recognize this calls into question the validity of E1 and K1 data.



Clearly, mining reliable resource data is crucial to the either the System Dynamics or StrikePRO methodology and also difficult to accomplish after the fact. NALDA's project officer made clear that NAVAIR recognizes data quality is an issue. They are working with TSD Orlando on an SBIR to address that issue. They documented several kinds of errors that they were aware existed in their databases that taken together called into question the validity of modeling using their data.

- Operational status codes conflicted, with ship fight hours logged while not deployed and deployed status but no ship flight hours logged in some cases.
- Poor documentation is affecting maintenance man-hours per flight hour calculations.
- Aircraft that were out of reporting status were being reported with flight hours against them.
- Aircraft listed at the wrong location.
- Incorrect "action taken code" used on 10% of cannibalization reports.

- Months where total flight hours were exceeded by the subset of total flight hours that represents strike training or air combat training.
- Months with no reported data at all.

NSAWC is currently very protective of their air wing performance data and has not yet cooperated in providing recent operational performance data. (When COMPACFLT pulled the funding on StrikePRO in mid-2000, NSAWC had provided in-depth operational performance data on sixteen air wings and BHA data on another eight air wings, but had not received any feedback on how well resource data allowed KAI to forecast an estimate of operational performance. Subsequent to that, it has been challenging to get NSAWC participation in outside analysis of their performance data.) However, if cooperation can be gained from both NSAWC and the Strike Fighter Weapons School, and NALDA is able to report some success in grooming Navy F/A-18 maintenance and SCIR databases, it may be possible to find corresponding performance and resource data from the most recent three years or so that do not introduce error.

8.2 Other Data related issues

It was hypothesized as part of the model development that simulator usage could have a positive effect on maintaining proficiency, however, it turns out that very few simulator hours are logged by active squadrons – the bulk of the hours are used by the Fleer Replacement Squadron (FRS) student pilots. Therefore, simulator usage was not considered in the demonstration. But with newer better fidelity simulators coming online and the potential for reduced flight hours or hours being unavailable due to aircraft maintenance issues, better data on simulator usage by fleet pilots needs to be collected.

Weapons expenditure data was not available at the squadron level. We were unable to determine when a squadron last performed an actual weapons drop. The initial theoretical model included a ten percent penalty against proficiency if a squadron had not dropped a weapon in 30 days. Data was not available to support this portion of the model and it was removed. Tracking of pilot/air crew weapons related activity is crucial resource data – and when coupled with the results of the weapons drop – hit or miss and CEP – adds significant information to analysts and planners.

The methodology for tracking and granting PMA points for high expense weapons events was not well defined. Some weapons, SLAM ER, JDAM, JSOW, HARM and MAVERICK, are singly assigned to a squadron maybe once per cycle. The squadron CO picks the pilot to launch the weapon. When that pilot departs the squadron, does proficiency in the special weapon depart as well?

There are opportunities for collection of performance data during the readiness cycle in addition to the performance at Fallon – SFARP, COMPTUEX, JTFEX, and actual operations. Collecting and archiving this performance data needs to be done in a formal and deliberate fashion to support analysis.

9.0 Lessons Learned

Socialization of the research approach with the participating activities and active sponsorship by the Commander Naval Air Forces (CNAF) was invaluable. Initially, some project time was expended in fostering coordination with CNAF staff. Subsequently, staff issues were fully resolved and the cooperation of the CNAF staff was superb for the duration of the experiment. Full cooperation from NSAWC was never realized.

Full resource data sets that match the time period of the operational performance data need to be assured before embarking on the analysis. Some data were not maintained after five years, or the data systems changed and the prior years data was not appropriately archived. Some useful data were either not

collected five years ago or were subsequently discarded. Identification of flight/mission/weapons data by pilot was the most obvious shortfall. This is because the nature of the performance data collected at NSAWC supports training and not performance evaluation. This is a subtle but challenging difference. Training by its very nature, if it to be effective, attempts to be non-threatening. Also, the training at Fallon is team-focused – not individual-based. These paradigms effect how data are collected, maintained and shared.

There are other opportunities for collection of individual and squadron performance data during the readiness cycle in addition to performance at Fallon – SFARP, COMPTUEX, JTFEX, and actual operations. Collecting and archiving this performance data needs to be done in a formal and deliberate fashion to support current and future analysis needs.

The Aviation Data Warehouse (ADW) is a treasure trove of information but it needs to evolve to better support the Naval Aviation Enterprise. Data validation between ADW, ACES and NALDA is required because of duplicative data entry/collection. Although the data are entered for different purposes, it is essentially the same data structured differently and entered by different people into different systems. ADW supports training readiness calculations, ACES supports flight hour program accounting and NALDA support aircraft maintenance.

10.0 Recommendations

It is recommended that the Naval Aviation Enterprise (NAE) take a close look at the requirements i.e., performance measures, methodology and culture associated with archiving linked sets of performance and resource data to enable current and future proficiency analyses. If the Navy leadership is serious about cost wise readiness, then some significant cultural practices in the way that individual performance data are collected, archived and managed will need major revision and in some cases, initial implementation.

Also, it is our recommendation that the Navy take positive steps to gather operational performance data for all NSAWC detachments, SFARP detachments, COMPTUEXes, JTFEXes and real world events. Using operational performance data drawn from all sources would enable a continuum of operational performance and help reveal a pattern of behavior. That can be used to better validate a system dynamics model for optimal patterns of resource allocation. The operational performance data should be included in the Aviation Data Warehouse (ADW) or some other designated single-source authoritative database.

Further use of the system dynamics method is recommended for analysis of training and readiness matrix requirements. Additional study in its use for Primary Mission Area analysis should be reconvened on resolution of the lack of sufficient performance measures and other data problems discussed in this report.

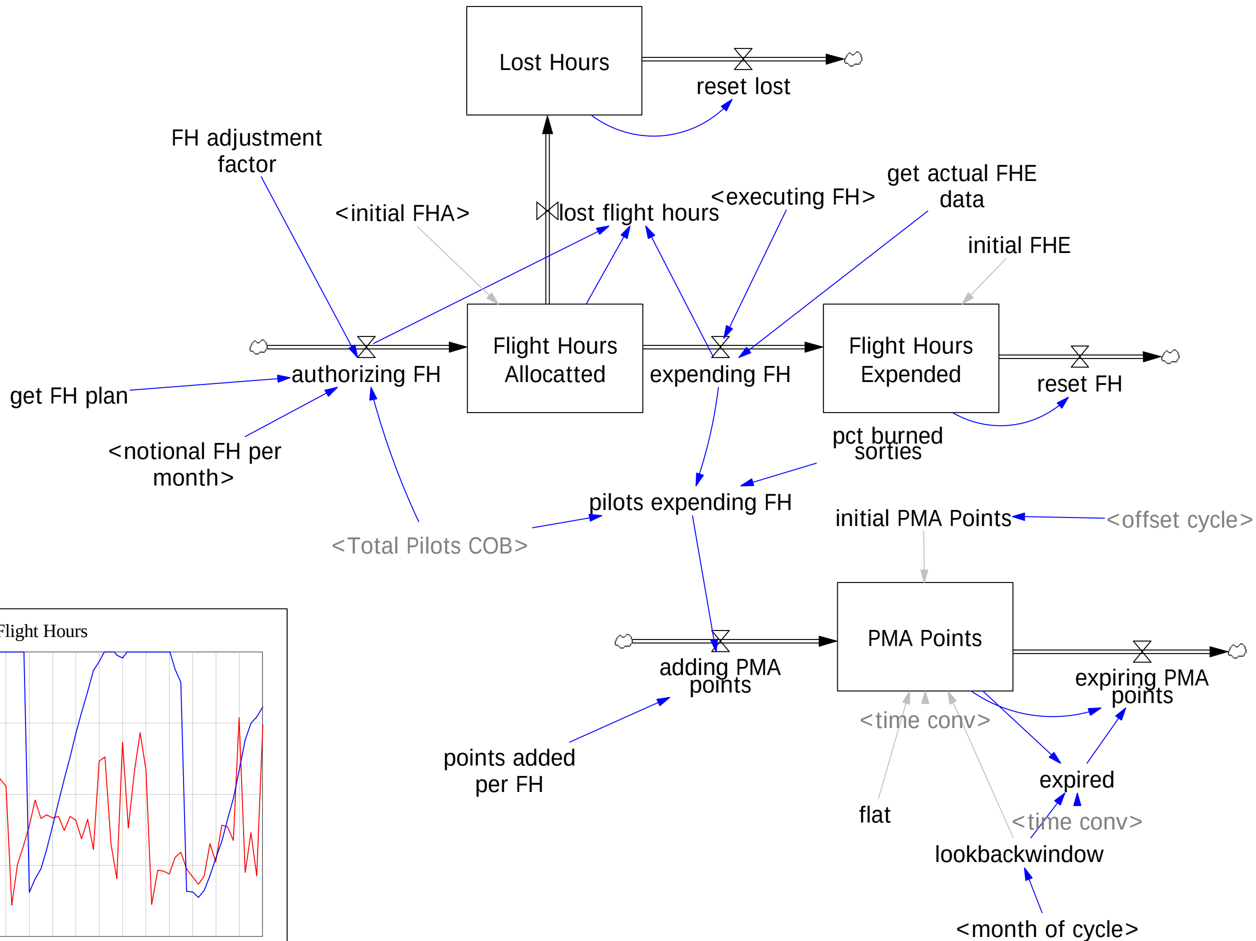
It is further recommended that the NAE perform a *complete systems level* data requirements analysis. This effort should include reviewing the legacy enterprise data systems for gaps in coverage, redundancy and duplicative data entry. Smart technology for aided data collection should be investigated at the pilot and aircraft levels. The analysis should identify all data that is collected and NOT collected, the processes for data collection, frequency, level of detail, data manipulation, aggregation, display, dissemination, data validation, integration of data systems storage, retrieval and archiving.

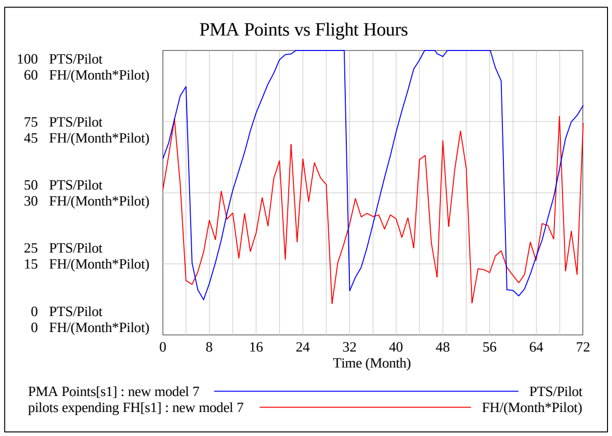
Recommended operational performance data that should be collected to support minimum performance based readiness analyses includes: Pilot identification, target data, event data, weapons data by type, date, time, hit evaluation data, amplifying event/mission data, flight data, other linked flight information, adversaries, wing and section data, environmental data.

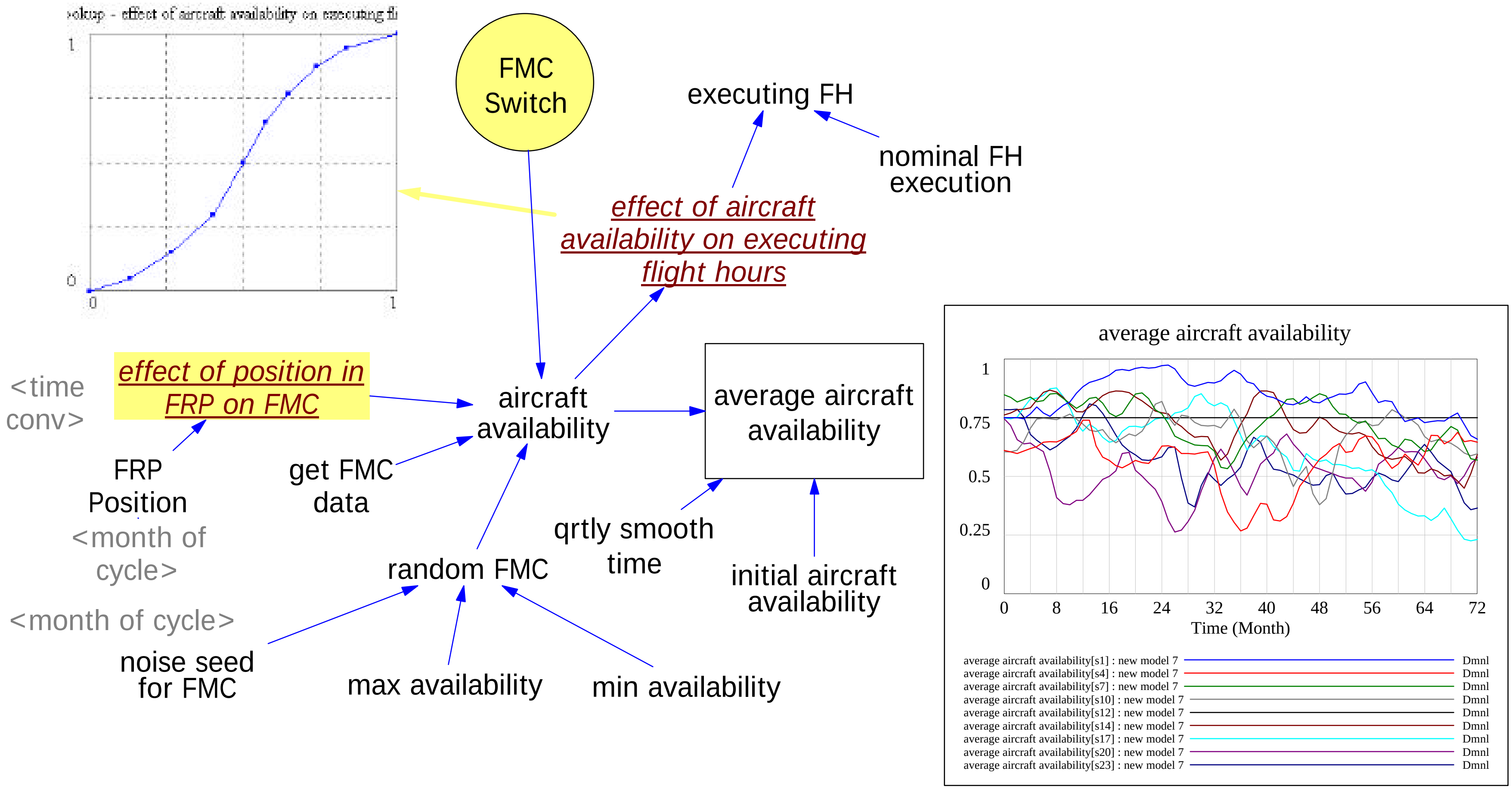
Appendix A

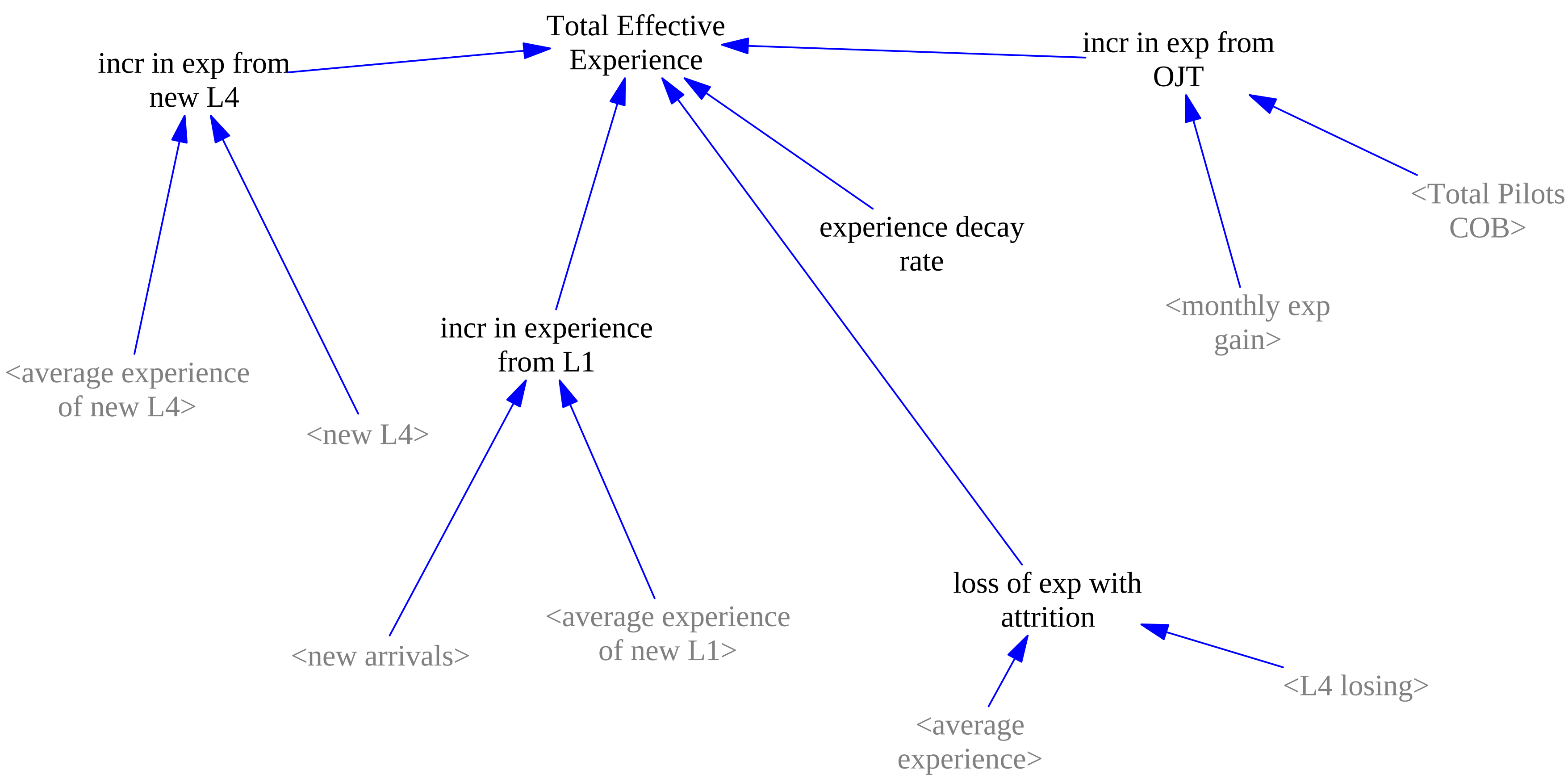
Aviation Readiness Production Model

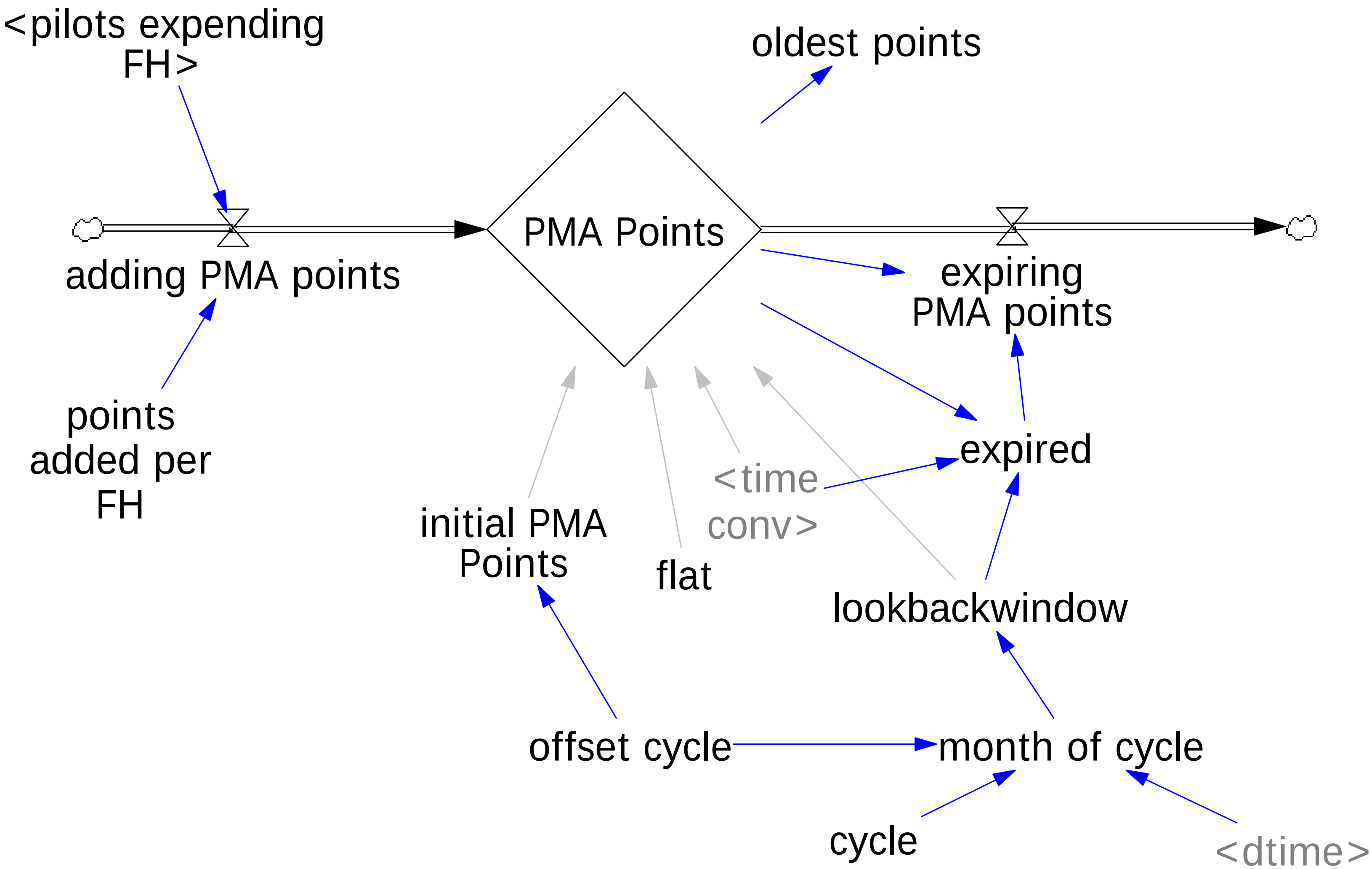
The following pages contain graphical model views followed by a complete variable dictionary and source code for the equation-based system dynamics model.

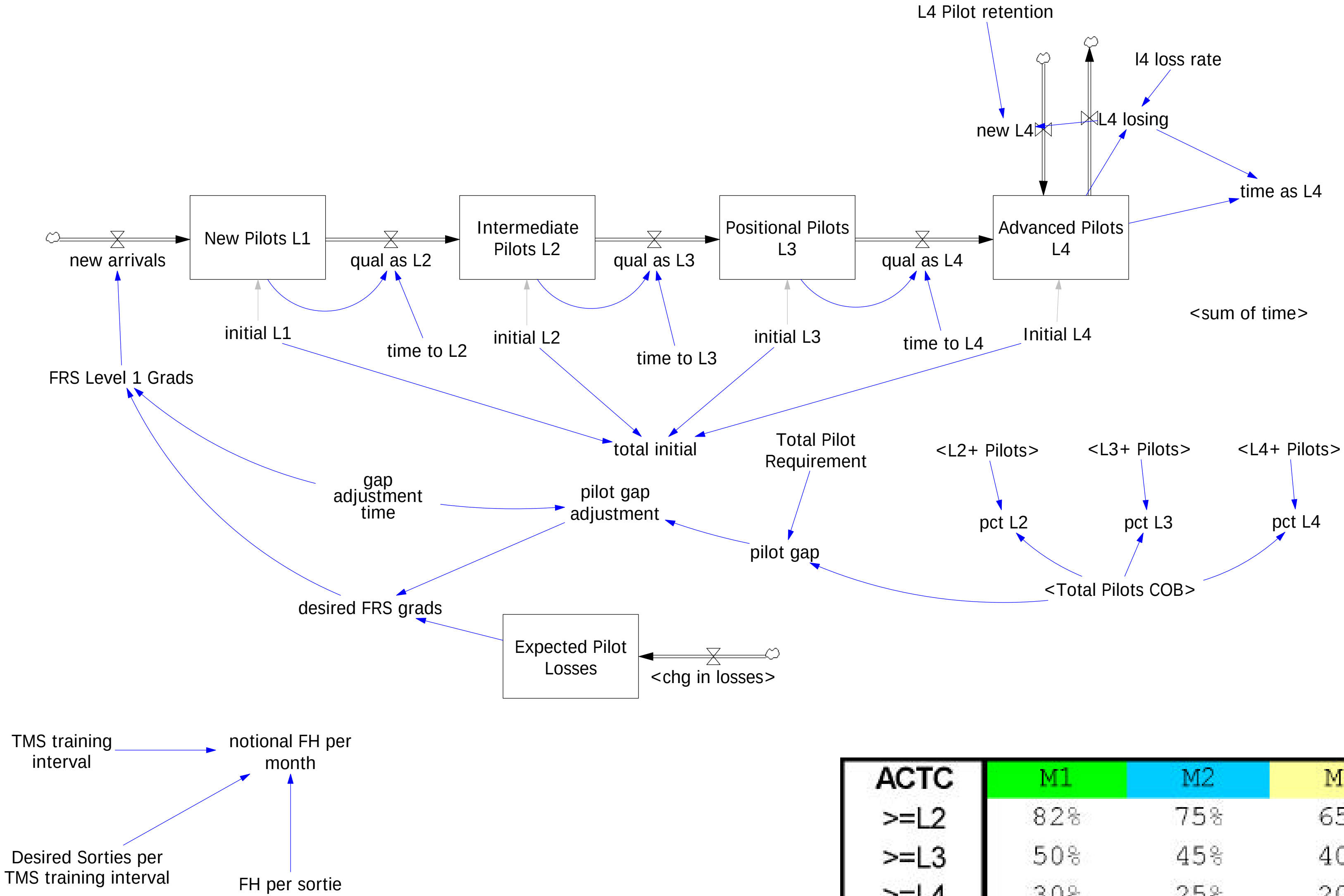

Lost Hours
reset lost
FH adjustment factor
<initial FHA>
lost flight hours
<executing FH>
get actual FHE data
initial FHE
Flight Hours Allocatted
Flight Hours Expended
authorizing FH
expending FH
reset FH
get FH plan
<notional FH per month>
pct burned sorties
pilots expending FH
<Total Pilots COB>
initial PMA Points
<offset cycle>
PMA Points
adding PMA points
expiring PMA points
<time conv>
points added per FH
expired
flat
<time conv>
lookbackwindow
<month of cycle>


PMA Points vs Flight Hours
100 PTS/Pilot
60 FH/(Month*Pilot)
75 PTS/Pilot
45 FH/(Month*Pilot)
50 PTS/Pilot
30 FH/(Month*Pilot)
25 PTS/Pilot
15 FH/(Month*Pilot)
0 PTS/Pilot
0 FH/(Month*Pilot)
0 8 16 24 32 40 48 56 64 72
Time (Month)
PMA Points[s1] : new model 7 PTS/Pilot
pilots expending FH[s1] : new model 7 FH/(Month*Pilot)

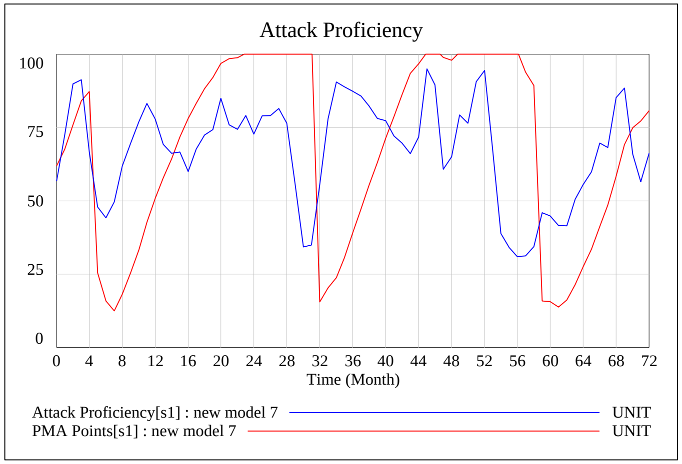

effect of aircraft availability on executing
1
0
0
1
FMC Switch
executing FH
nominal FH execution
effect of aircraft availability on executing flight hours
<time conv>
effect of position in FRP on FMC
aircraft availability
average aircraft availability
FRP Position
get FMC data
<month of cycle>
qrtly smooth time
random FMC
initial aircraft availability
<month of cycle>
noise seed for FMC
max availability
min availability
average aircraft availability
1
0.75
0.5
0.25
0
0
8
16
24
32
40
48
56
64
72
Time (Month)
average aircraft availability[s1] : new model 7 Dmnl
average aircraft availability[s4] : new model 7 Dmnl
average aircraft availability[s7] : new model 7 Dmnl
average aircraft availability[s10] : new model 7 Dmnl
average aircraft availability[s12] : new model 7 Dmnl
average aircraft availability[s14] : new model 7 Dmnl
average aircraft availability[s17] : new model 7 Dmnl
average aircraft availability[s20] : new model 7 Dmnl
average aircraft availability[s23] : new model 7 Dmnl


Total Effective Experience
incr in exp from new L4
incr in exp from OJT
experience decay rate
<Total Pilots COB>
<monthly exp gain>
incr in experience from L1
<average experience of new L4>
<new L4>
loss of exp with attrition
<new arrivals>
<average experience of new L1>
<L4 losing>
<average experience>

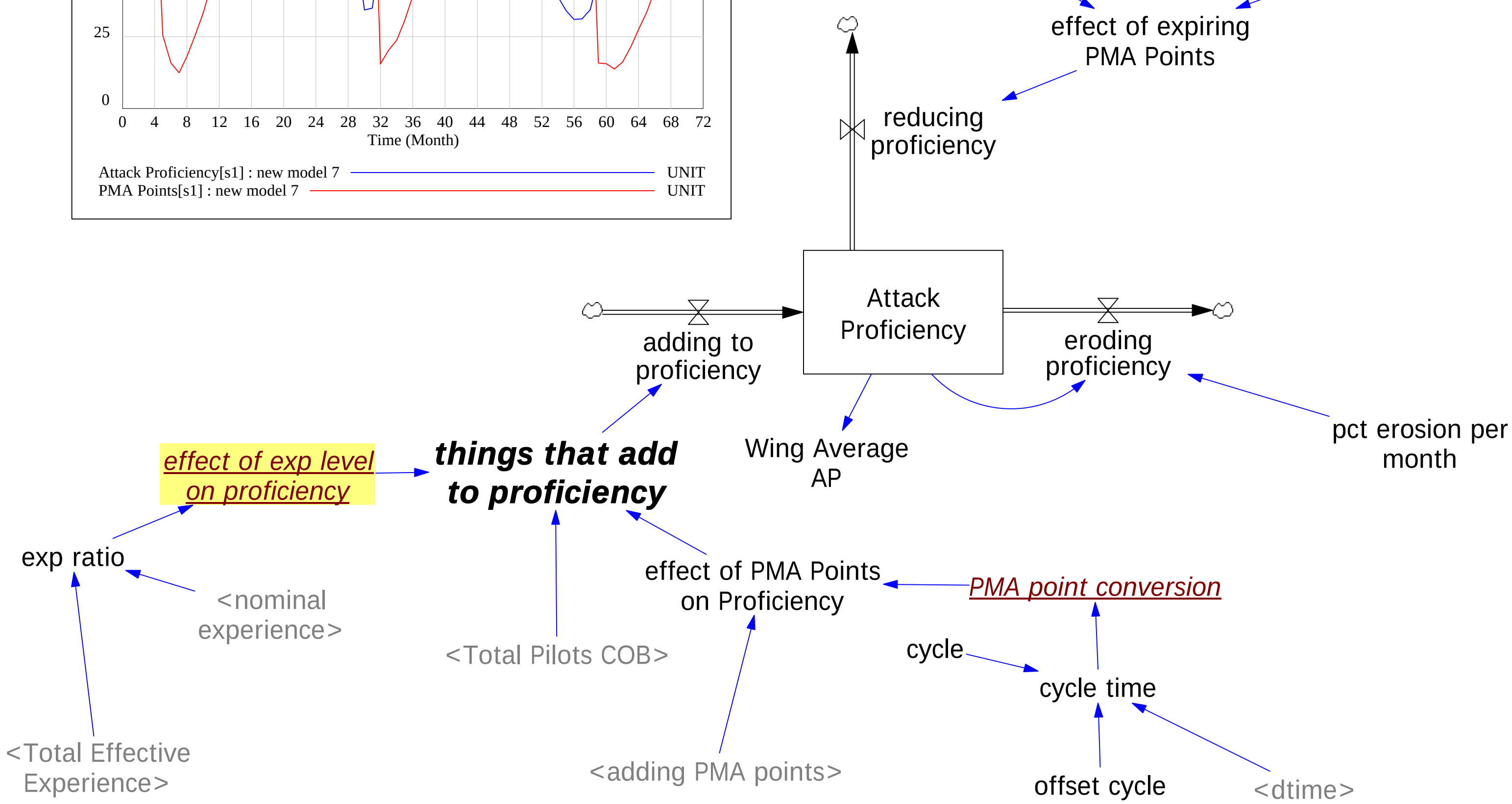

<pilots expending FH>
oldest points
PMA Points
adding PMA points
expiring PMA points
points added per FH
expired
<time conv>
initial PMA Points
flat
lookbackwindow
offset cycle
month of cycle
cycle
<dtime>

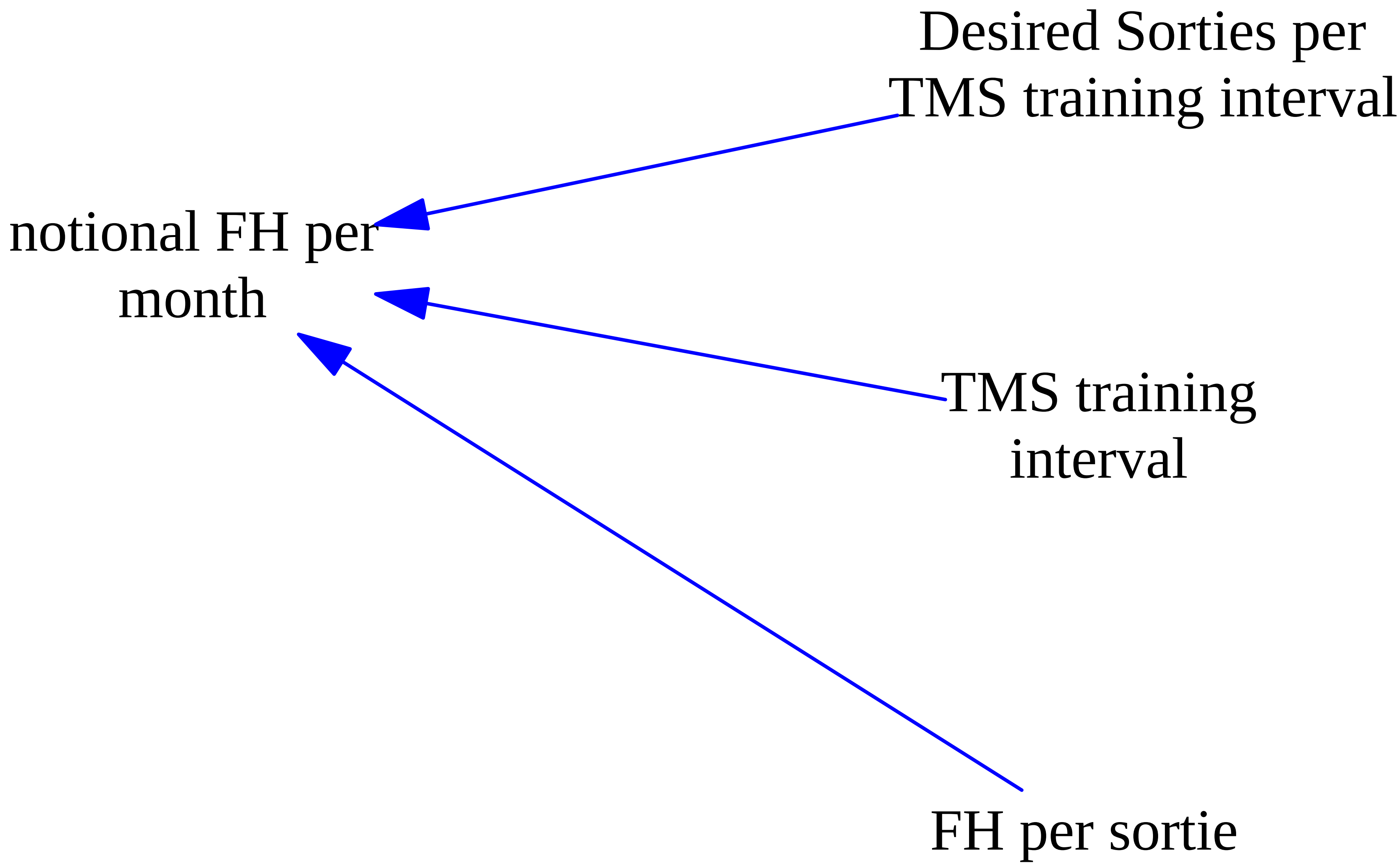


ACTC	M1	M2	M3
>=L2	82%	75%	65%
>=L3	50%	45%	40%
>=L4	30%	25%	20%



"Operationally proficient/competent crews equate to operationally (combat) ready crews." CNAFINST 3500.1B




Desired Sorties per
TMS training interval
notional FH per
month
TMS training
interval
FH per sortie

Aviation Readiness Production Model
Flighthoursv7.mdl dated 8/12/2005
System Dynamics Model
Variable Dictionary

****
Key:
model variable =
equation expressed in FUNCTIONS of variables or constant value
~ units
~ comments/remarks
****
.flighthoursV7
***~
Developed by CACI Dynamic Systems Inc. under contract to the Office of Naval Research (ONR). POC is Mike McDevitt, mmcdevitt@caci.com
Phone: 858-695-8220 x1457
|

adding PMA points[SQDN]=
IF THEN ELSE (Time >switch time, pilots expending FH[SQDN] * points added per FH, get actual FHE data\
[SQDN] * points added per FH/pilots expending FH[SQDN])
~ PTS/(Month * Pilot)
~ |

adding to proficiency[SQDN]=
things that add to proficiency[SQDN]
~ UNITS/Month
~ |

Advanced Pilots L4[SQDN]= INTEG (
qual as L4[SQDN]-L4 losing[SQDN] + new L4[SQDN],
Initial L4[SQDN])
~ Pilots
~ |

after points=
0.35
~ PTS/FH
~ |

aircraft availability[SQDN]=

IF THEN ELSE(FMC Switch = 0, random FMC* effect of position in FRP on FMC[SQDN], get FMC data\
)
~ Dmnl
~ Get Aircraft FMC data from CNAF/NALCOMIS
|

annual losses[SQDN]=
Expected Pilot Losses[SQDN] * months per year[SQDN]
~ Pilot
~ Get historical loss data from TRMS
~ :SUPPLEMENTARY
|

Attack Proficiency[SQDN]= INTEG (
adding to proficiency[SQDN]-eroding proficiency[SQDN] - reducing proficiency[SQDN],
initial proficiency[SQDN])
~ UNIT
~ |

authorizing FH[SQDN]=
get FH plan[SQDN] * notional FH per month[SQDN] * Total Pilots COB[SQDN] * FH adjustment factor
~ FH/Month
~ Data from ACES read in here: validate assumptions
|

average aircraft availability[SQDN]=
SMOOTHI(aircraft availability[SQDN], qrtly smooth time, initial aircraft availability\
[SQDN])
~ Dmnl
~ Quarterly Smooth of Stochastic Availability
~ :SUPPLEMENTARY
|

average experience[SQDN]=
Total Effective Experience[SQDN]/Total Pilots COB[SQDN]
~ Month/Pilot
~ |

average experience of new L1[SQDN]=
1
~ Month/Pilot

~ |

average experience of new L4[SQDN]=
36
~ Month/Pilot
~ |

Break time=
48
~ Month
~ 48
|

chg in losses[SQDN]=
((L4 losing[SQDN]-new L4[SQDN]) - Expected Pilot Losses[SQDN])/loss smoothing time
~ Pilots/(Month * Month)
~ |

comment1=1
~ Dmnl
~ 85% of PRMAR point value for a given PRMAR = Operationally Proficient
70% of PRMAR point value for a given PRMAR = Operationally Competent
55% of PRMAR point value for a given PRMAR = Operationally Safe
~ :SUPPLEMENTARY
|

comment2=1
~ Dmnl
~ M-1 = >=85% of crews on board have >=70% of T&R points in that PRMAR
M-2 = >=70% of crews on board have >=70% of T&R points in that PRMAR
M-3 = >=55% of crews on board have >=70% of T&R points in that PRMAR
M-4 = < 55% of crews on board have >=70% of T&R points in that PRMAR
~ :SUPPLEMENTARY
|

comment3 = 1
~ Dmnl
~ In order to calculate the number of operationally (combat) ready crews,

the individual aircrew readiness must be tied into a crew (as defined by each \
T/M/S).
Crew PRMAR points are based on the PRMAR points achieved and divided by the number of
available (COB) crews for that PRMAR. SHARP will track individual PRMAR points \
achieved
and match crews together based on crew positions, ACTC requirements, crew \
qualification
requirements and aircrew availability. SHARP will then calculate the optimum number \
of
operationally (combat) ready crews.
~ :SUPPLEMENTARY
|

comment4=0
~ Dmnl
~ LEVEL QUALIFICATION/DESIGNATION
L1 (I) FRS Complete
L2 (II) Intermediate (e.g., Combat Wingman)
L3 (III) Positional (e.g., Combat Section Lead, ASW TACCO)
L4 (IV) Advanced (e.g., Combat Division Lead, Mission Commander)
L5 (V) Weapons and Tactics Instructor (WTI)
~ :SUPPLEMENTARY
|

comment5=
0
~ Dmnl
~ FRS Completion Entry Level Readiness Measurement. Type Wing Commanders \
will standardize the T&R points an FRS graduate receives in each PRMAR. \
T&R points will be determined based upon the assessed core competency of \
FRS CAT I through V aircrew upon syllabus completion. The PRMAR point \
baseline should be viewed as an FRS diploma that carries the graduate \
through a grace period starting with the first Fleet squadron day and \
terminating at the completion of one T/M/S defined training interval. At \
the end of the first training interval, the baseline will be removed. The \

FRS baseline is defined by PRMAR points.
~ :SUPPLEMENTARY
|

cycle[SQDN]=
27
~ Dmnl
~ FRP cycle for CV Wings is 27 Months based on CVN Maintenance
|

cycle time[SQDN]=
(MODULO(ABS(+ dtime + offset cycle[SQDN]), cycle[SQDN])/cycle[SQDN])
~ Dmnl
~ add a shift to dtime of the depolyment date to reset the counter by \
squadron.
|

desired FRS grads[SQDN]=
Expected Pilot Losses[SQDN] + pilot gap adjustment[SQDN]
~ Pilots/Month
~ |

Desired Sorties per TMS training interval[SQDN]=
54
~ Sorties
~ |

dtime=
Time * time conv
~ Dmnl
~ |

effect of aircraft availability on executing flight hours[SQDN]= WITH LOOKUP (
aircraft availability[SQDN],
([(0,0)-
(1,1)],(0,0),(0.134557,0.0482456),(0.269113,0.149123),(0.40367,0.29386),(0.501529\

,0.5),(0.571865,0.653509),(0.648318,0.763158),(0.740061,0.872807),(0.834862,0
.942982\
),(1,1),(20,1)))
~ Dmnl
~ Some flights can be conducted with NMC aircraft up to a point. Double \
cycling of aircraft up to some limit, then a sharp drop off in \
availability should occur. Place holder\!\!\!
|

```
effect of exp level on proficiency[SQDN]= WITH LOOKUP (
	exp ratio[SQDN],
		([(0,0.8)-
(2,1.5)],(0,0.85),(0.25,0.85),(0.45,0.85),(0.611621,0.867544),(0.721713,0.907456\
		),(0.82,0.95),(1,1),(1.24159,1.03333),(1.75,1.05),(2,1.05) ))
	~	Dmnl
	~	8/3/2/2  nominal design ->> 5/2/4/4 worst case - no degradation above \
		nominal no more than a 15% drop at worst case\!\!\!
	|

effect of expiring PMA Points[SQDN]=
	expiring PMA points[SQDN] * pct per pilot
	~	UNITS/Month
	~		|

effect of maint cycle=
	1
	~	Dmnl
	~		~	:SUPPLEMENTARY
	|

effect of NCEA availability on proficiency[SQDN]=
	1 * loss of weapons qual with transfer[SQDN]
	~	Dmnl
	~	placeholder
	|

effect of PMA Points on Proficiency[SQDN]=
	adding PMA points[SQDN] * PMA point conversion[SQDN]
	~	UNITS/(Month * Pilot)
	~		|

effect of position in FRP on FMC[SQDN]= WITH LOOKUP (
	month of cycle[SQDN] * time conv,
		([(-8,0)-(27,2)],(-
6,1),(0,1),(1,1),(2,1),(3,1),(4,1),(5,0.85),(6,0.85),(7,0.85),(8\
		,0.85),(10,0.85),(12,1),(14,1.15),(18,1.15),(20,1.15),(24,1.15),(26,1),(27,1)
))
	~	Dmnl
	~		|

effect of simulator usage on proficiency[SQDN]=
	1
	~	Dmnl
```

~ placeholder
|

eroding proficiency[SQDN]=
Attack Proficiency[SQDN]/proficiency erosion time
~ UNITS/Month
~ |

executing FH[SQDN]=
effect of aircraft availability on executing flight hours[SQDN] * nominal FH execution\
[SQDN]
~ Dmnl/Month
~ 100 percent utilization of the plan is unlikely - it will probably be \
something less on average due to constraints like weather and equipment \
failures. Sometimes a period of surge over 100% will occur to utilize more \
FH to correct a training gap and recover the training schedule. This \
should be a dynamic variable based on constraints and goals. It could be \
a data steam from ACES or ADW.
|

exp ratio[SQDN]=
Total Effective Experience[SQDN]/nominal experience
~ Dmnl
~ |

Expected Pilot Losses[SQDN]= INTEG (
chg in losses[SQDN],
0.4)
~ Pilots/Month
~ |

expending FH[SQDN]=
Flight Hours Allocatted[SQDN]* executing FH[SQDN]
~ FH/Month
~ |

experience decay rate[SQDN]=
0
~ 1
~ notionally set to zero because of T&R matrix.
|

expired[SQDN]=

QUEUE AGE IN RANGE(PMA Points[SQDN],lookbackwindow[SQDN],500)
~ PTS/Pilot
~ |

expiring PMA points[SQDN]=
MIN(PMA Points[SQDN], expired[SQDN])/TIME STEP
~ PTS/(Month*Pilot)
~ |

FH adjustment factor=
1.5
~ Dmnl
~ |

FH per sortie[SQDN]=
1.5
~ FH/Sortie/Pilot
~ |

flat(
[(0,0)-(3,2)],(0,1),(3,1))
~ Dmnl
~ |

Flight Hours Allocatted[SQDN]= INTEG (
authorizing FH[SQDN]-expending FH[SQDN] -lost flight hours[SQDN],
initial FHA[SQDN])
~ FH
~ |

Flight Hours Expended[SQDN]= INTEG (
expending FH[SQDN] -reset FH[SQDN],
initial FHE[SQDN])
~ FH
~ |

FMC Switch=
1
~ Dmnl
~ set to other than zero to get data from spreadsheet.
|

FRS Level 1 Grads[SQDN]=
INTEGER(PULSE TRAIN(0, TIME STEP, gap adjustment time[SQDN],
FINAL TIME) * desired FRS grads\

[SQDN] * gap adjustment time[SQDN])
~ Pilots
~ Pulse so many pilots per quarter into the squadron from the FRS at L1
|

gap adjustment time[SQDN]=
4
~ Month
~ |

get actual FHE data[SQDN]:=
GET XLS DATA('nominal flight hours.xls' , 'ACES_Data' , 'B' , 'C6')
~ FH/Month
~ |

get FH plan[SQDN]:=
GET XLS DATA('nominal flight hours.xls' , 'Sheet4' , 'B' , 'I2')
~ Dmnl
~ |

get FMC data:=
GET XLS DATA('nominal flight hours.xls' , 'AMRR_Data' , 'B' , 'C6')
~ Dmnl
~ |

incr in exp from new L4[SQDN]=
new L4[SQDN] * average experience of new L4[SQDN]
~ 1
~ |

incr in exp from OJT[SQDN]=
Total Pilots COB[SQDN] * monthly exp gain[SQDN]
~ 1
~ |

incr in experience from L1[SQDN]=
average experience of new L1[SQDN] * new arrivals[SQDN]
~ 1
~ |

initial aircraft availability[SQDN]:=
GET XLS DATA('nominal flight hours.xls' , 'AMRR_data' , 'B' , 'C6')
~ Dmnl
~ |

initial FHA[SQDN]=
notional FH per month[SQDN] * Total Pilots COB[SQDN] /time conv * initial aircraft availability\
[SQDN]
~ FH
~ |

initial FHE[SQDN]=
15
~ FH
~ |

initial L1[SQDN]=
2
~ Pilots
~ Read in by squadron and type.
|

initial L2[SQDN]=
2
~ Pilots
~ |

initial L3[SQDN]=
3
~ Pilots
~ |

Initial L4[SQDN]=
8
~ Pilots
~ |

initial PMA Points[SQDN]= WITH LOOKUP (
offset cycle[SQDN],
([(0,0)-(36,100)],(0,25),(10.789,28.9474),(16.5138,36.8421),(19.2661,50),(21.578,58.7719\
),(24.3303,64.9123),(27,65)))
~ PTS/Pilot
~ |

initial proficiency[SQDN]=
initial PMA Points[SQDN] * (1-percent degradation) * unit pt conversion
~ UNIT
~ |

Intermediate Pilots L2[SQDN]= INTEG (
qual as L2[SQDN]-qual as L3[SQDN],
initial L2[SQDN])
~ Pilots
~ |

"L2+ Pilots"[SQDN]=
Intermediate Pilots L2[SQDN]+"L3+ Pilots"[SQDN]
~ Pilots
~ |

"L3+ Pilots"[SQDN]=
"L4+ Pilots"[SQDN]+ Positional Pilots L3[SQDN]
~ Pilots
~ |

L4 losing[SQDN]=
Advanced Pilots L4[SQDN]* l4 loss rate[SQDN]
~ Pilots/Month
~ |

l4 loss rate[SQDN]=
0.25
~ 1/Month
~ lose 3 L4s per year or one per quarter on average
|

L4 Pilot retention[SQDN]=
0.75
~ Dmnl
~ |

"L4+ Pilots"[SQDN]=
Advanced Pilots L4[SQDN]
~ Pilots
~ |

lookbackwindow[SQDN]=
MIN(IF THEN ELSE(month of cycle[SQDN] > 3, month of cycle[SQDN], 3), 12)
~ Month
~ |

loss of exp with attrition[SQDN]=

average experience[SQDN] * L4 losing[SQDN]
~ 1
~ |

loss of weapons qual with transfer[SQDN]=
1
~ Dmnl
~ placeholder
|

loss smoothing time=
3
~ Month
~ |

lost flight hours[SQDN]=
IF THEN ELSE(MODULO(Time, 6) = 1, Flight Hours Allocatted[SQDN]/TIME STEP - expending FH\
[SQDN]- authorizing FH[SQDN], 0)
~ FH/Month
~ |

Lost Hours[SQDN]= INTEG (
lost flight hours[SQDN] -reset lost[SQDN],
0)
~ FH
~ |

max availability=
0.9
~ Dmnl
~ 90% Maximum Aircraft Availability for 10 ac sqdn 11/12 for 12 ac sqdn
|

min availability=
0.5
~ Dmnl
~ 50% minimum set as default.
|

month of cycle[SQDN]=
ABS(MODULO(dtime+ offset cycle[SQDN], cycle[SQDN]))
~ Dmnl
~ |

monthly exp gain[SQDN]=
1
~ Month/(Pilot* Month)
~ |

months per year[SQDN]=
12
~ Month
~ |

new arrivals[SQDN]=
FRS Level 1 Grads[SQDN]/TIME STEP
~ Pilots/Month
~ |

new L4[SQDN]= DELAY FIXED (
L4 losing[SQDN]* L4 Pilot retention[SQDN], 3, L4 losing[SQDN])
~ Pilots/Month
~ |

New Pilots L1[SQDN]= INTEG (
new arrivals[SQDN]-qual as L2[SQDN],
initial L1[SQDN])
~ Pilots
~ |

noise seed for FMC=
999
~ Dmnl
~ |

nominal experience=
600
~ Month
~ 36 months flying time * 15 Pilots = 510 months of expereince
|

nominal FH execution[SQDN]=
1
~ Dmnl/Month
~ Normally 1.0 but reduced by 2% due to weather/fog - see lost sorties
|

notional FH per month[SQDN]=

Desired Sorties per TMS training interval[SQDN] * FH per sortie[SQDN]/TMS training interval\
[SQDN]
~ FH/Month/Pilot
~ Flight Hours per pilot per month for a 90 day training interval.
|

offset cycle[SQDN]=
GET XLS CONSTANTS('nominal flight hours.xls' , 'Sheet4' , 'AI2')
~ Dmnl
~ |

oldest points[SQDN]=
QUEUE AGE OLDEST(PMA Points[SQDN])
~ Month
~ ~ :SUPPLEMENTARY
|

pct burned sorties=
0.04
~ Dmnl
~ 1 0f 25 lost mission sorties
|

pct L2[SQDN]=
"L2+ Pilots"[SQDN]/Total Pilots COB[SQDN]
~ 1
~ ~ :SUPPLEMENTARY
|

pct L3[SQDN]=
"L3+ Pilots"[SQDN]/Total Pilots COB[SQDN]
~ 1
~ ~ :SUPPLEMENTARY
|

pct L4[SQDN]=
"L4+ Pilots"[SQDN]/Total Pilots COB[SQDN]
~ 1
~ ~ :SUPPLEMENTARY
|

pct per pilot=
0.055
~ (UNIT* Pilot)/PTS

~ 1/15 = .067
|

percent degradation=
0.083
~ Dmnl
~ 1/12 = .083   1/10 = 0.1
|

pilot gap[SQDN]=
Total Pilot Requirement[SQDN]-Total Pilots COB[SQDN]
~ Pilots
~ |

pilot gap adjustment[SQDN]=
pilot gap[SQDN]/gap adjustment time[SQDN]
~ Pilots/Month
~ |

pilots expending FH[SQDN]=
(1-pct burned sorties) * expending FH[SQDN]/Total Pilots COB[SQDN]
~ FH/(Month*Pilot)
~ |

PMA point conversion[SQDN]= WITH LOOKUP (
cycle time[SQDN],
([(0,0)-(1,1)],(0,0.925),(1,0.3)))
~ UNITS/PTS
~ More or less points depending upon where in the cycle the points are \
accrued - should be dynamic based on learning curve function.OLD: \
([(0,0)-
(1,1)],(0,0.1),(0.122324,0.118421),(0.238532,0.166667),(0.324159,0.\

214912),(0.394495,0.285088),(0.431193,0.359649),(0.501529,0.495614),(0.5963\
3,0.596491),(0.712538,0.649123),(0.856269,0.675439),(1,0.67))\!\!\!
|

PMA Points[SQDN]=
QUEUE FIFO(adding PMA points[SQDN], expiring PMA points[SQDN], flat,
initial PMA Points\
[SQDN], lookbackwindow[SQDN])
~ PTS/Pilot
~ |

points added per FH=

IF THEN ELSE (Time < Break time, prebreak points, after points)
~ PTS/FH
~ Variable depending on lots of things - place holder for now. Place in \
Cycle, number of events - whether repetitive training or not. Type of \
training. Could be broken down by PMA type - Airmanship, Bombing, Fighter \
or other (Maint or demo flights). Adjuted down from .3 based on better fit \
of the data.
|

Positional Pilots L3[SQDN]= INTEG (
qual as L3[SQDN]-qual as L4[SQDN],
initial L3[SQDN])
~ Pilots
~ |

prebreak points=
0.25
~ PTS/FH
~ |

proficiency erosion time=
1
~ Month
~ |

qrtly smooth time=
3
~ Month
~ 3 months per quarter
|

qual as L2[SQDN]=
New Pilots L1[SQDN]/time to L2
~ Pilots/Month
~ |

qual as L3[SQDN]=
Intermediate Pilots L2[SQDN]/time to L3
~ Pilots/Month
~ |

qual as L4[SQDN]=
Positional Pilots L3[SQDN]/time to L4

~ Pilots/Month
~ |

random FMC=
RANDOM UNIFORM(min availability, max availability, noise seed for FMC)
~ Dmnl
~ |

reducing proficiency[SQDN]=
things that reduce proficiency[SQDN] + effect of expiring PMA Points[SQDN]
~ UNITS/Month
~ |

reset FH[SQDN]=
IF THEN ELSE(Time >1, MAX(IF THEN ELSE(MODULO(Time, 12) = 1, Flight Hours Expended[\
SQDN]/TIME STEP, 0), 0) , 0
)
~ FH/Month
~ Used for annual reset of hours
|

reset lost[SQDN]=
MAX(IF THEN ELSE(MODULO(Time, 12) = 1, Lost Hours[SQDN]/TIME STEP, 0), 0)
~ FH/Month
~ |

SQDN:
s1, s2, s3, s4, s5, s6, s7, s8, s9, s10,
s11, s12,s13,s14,s15,s16,s17,s18,s19,
s20,s21,s22,s23,s24,s25
~
~ Place holder - VFA squadron data will be read in from a spreadsheet by \
squadron. Squadrons will be associated by wing and grouped by subscript - \
for example: w1: s1, s2, s3; w2: s4, s5, s6
|

stw mission late switch=
0
~ Dmnl
~ stw mission late switch turns on and off. 0 is off 1 is on.
|

sum of time[SQDN]=
time as L4[SQDN] + time to L2 + time to L3 + time to L4
~ Month
~ ~ :SUPPLEMENTARY
|

switch time=
-1
~ Month
~ |

things that add to proficiency[SQDN]=
1 * effect of PMA Points on Proficiency[SQDN]
* Total Pilots COB[SQDN]
* effect of simulator usage on proficiency[SQDN]
* effect of NCEA availability on proficiency[SQDN]
* effect of exp level on proficiency[SQDN]
~ UNITS/Month
~ Placeholder + PMA Ponts NCEA and Simulator time needs to be added as \
well experience effect
|

things that reduce proficiency[SQDN]=
IF THEN ELSE (time since last STW Mission[SQDN] = 1.25, Attack Proficiency[SQDN] * percent degradation\
/TIME STEP, 0) * stw mission late switch
~ UNITS/Month
~ Loss of proficeincy of 10% if more than 30 days has traspired since last \
bomb drop activity. May not be historical data available to poulate this \
variable. stw mission late switch turns on and off. 0 is off 1 is on.
|

time as L4[SQDN]=
Advanced Pilots L4[SQDN]/L4 losing[SQDN]
~ Month
~ |

time conv=
1
~ 1/Month
~ Used to create dimensionless time for lookup tables
|

time since last STW Mission[SQDN]= WITH LOOKUP (

```
        dtime,
                ([(0,0)-
(100,4)],(0,0),(7,2),(12,0),(17,0),(20,1),(23.8532,0),(34,1),(38,0),(49,1),\
                (57,0),(64.526,1),(72,0),(74,1),(87,0),(88,1),(100,0) ))
        ~       Dmnl
        ~       sample - needs to be dynamic based on data\!\!\!
        |

time to L2=
        6
        ~       Month
        ~       100% empty in 6 Months plus
        |

time to L3=
        12
        ~       Month
        ~       12 months for 100% empty
        |

time to L4=
        12
        ~       Month
        ~       12 Months
        |

TMS training interval[SQDN]=
        3
        ~       Month
        ~       90 days or 3 Months
        |

Total Effective Experience[SQDN]= INTEG (
        incr in experience from L1[SQDN] + incr in exp from OJT[SQDN] - loss of exp
with attrition\
                [SQDN] - experience decay rate[SQDN] + incr in exp from new
L4[SQDN],
                570)
        ~       Month
        ~       15 pilots with an average of 2 years experience
        |

total initial[SQDN]=
        initial L1[SQDN] + initial L2[SQDN] + initial L3[SQDN] + Initial L4[SQDN]
        ~       Pilot
```

~ ~ :SUPPLEMENTARY
|

Total Pilot Requirement[SQDN]=
15
~ Pilots
~ Read in by squadron and type - For example 15 pilots for F/A-18C and \
F/A-18E squadrons with a PAA of ten aircraft. F/A-18F and F-14
squadrons \
have different manning requirements. Manning may have changed over
time - \
used to be 17 pilots. Make time series data read in from XLS
|

Total Pilots COB[SQDN]=
Advanced Pilots L4[SQDN]+Intermediate Pilots L2[SQDN]+New Pilots
L1[SQDN]+Positional Pilots L3\
[SQDN]
~ Pilots
~ |

unit pt conversion=
1
~ UNIT*Pilot/PTS
~ |

w1 : s1, s2, s3
~
~ First CVW
|

w2 : s4, s5, s6
~
~ 2nd CVW
|

w3:
s7, s8, s9
~
~ |

w4: s10, s11
~
~ |

w5: s12, s13
~
~ |

w6: s14, s15, s16
~
~ |

w7: s17, s18, s19
~
~ |

w8: s20, s21, s22
~
~ |

w9: s23, s24, s25
~
~ |

Wing Average AP[w1]=
(Attack Proficiency[s1]+ Attack Proficiency[s2] + Attack Proficiency[s3])/ELMCOUNT(w1\
) ~~|
Wing Average AP[w2]=
(Attack Proficiency[s4]+ Attack Proficiency[s5] + Attack Proficiency[s6])/ELMCOUNT(w2\
) ~~|
Wing Average AP[w3]=
(Attack Proficiency[s7]+ Attack Proficiency[s8] + Attack Proficiency[s9])/ELMCOUNT(w3\
) ~~|
Wing Average AP[w4]=
(Attack Proficiency[s10]+ Attack Proficiency[s11])/ELMCOUNT(w4) ~~|
Wing Average AP[w5]=
(Attack Proficiency[s12]+ Attack Proficiency[s13])/ELMCOUNT(w5) ~~|
Wing Average AP[w6]=
(Attack Proficiency[s14]+ Attack Proficiency[s15] + Attack Proficiency[s16])/ELMCOUNT\
(w6) ~~|
Wing Average AP[w7]=
(Attack Proficiency[s17]+ Attack Proficiency[s18] + Attack Proficiency[s19])/ELMCOUNT\
(w7) ~~|
Wing Average AP[w8]=

(Attack Proficiency[s20]+ Attack Proficiency[s21] + Attack Proficiency[s22])/ELMCOUNT\
(w8) ~~|
Wing Average AP[w9]=
(Attack Proficiency[s23]+ Attack Proficiency[s24] + Attack Proficiency[s25])/ELMCOUNT\
(w9)
~ UNIT
~ ~ :SUPPLEMENTARY
|

****
.Control
**~
Simulation Control Parameters
|

FINAL TIME = 72
~ Month
~ The final time for the simulation.
|

INITIAL TIME = 0
~ Month
~ The initial time for the simulation.
|

SAVEPER = 1
~ Month [0,60]
~ The frequency with which output is stored.
|

TIME STEP = 0.25
~ Month [0,60]
~ The time step for the simulation.
||

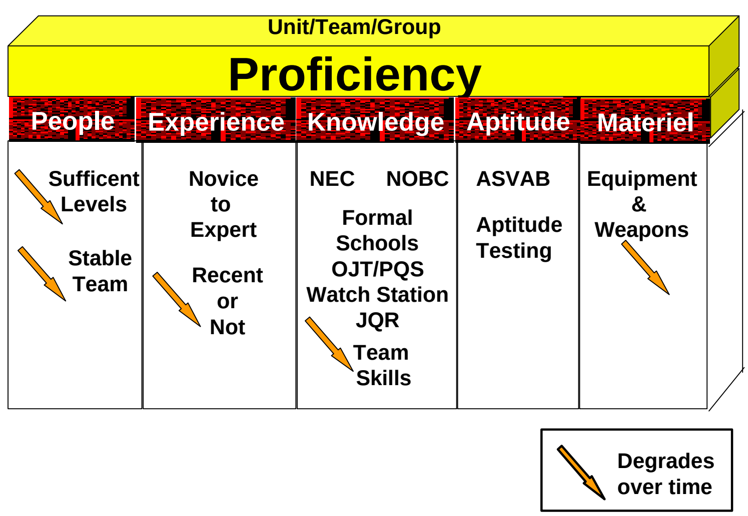
The StrikePRO Benchmark

Background

In 1997, Kapos Associates, Inc. (KAI) began to explore possible cause-and-effect relationships between ASW resources and ASW proficiency in a COMPACFLT-funded project known as the ASWPRO model. In 1999, the PRO methodology was applied to the area of aviation strike warfare in another COMPACFLT-funded project known as the StrikePRO model. Mr. Michael McDevitt pioneered that work, overseeing both projects until leaving KAI late in 2000. By that time, ASWPRO was reaching maturity as an experiment and StrikePRO was still in its infancy. COMPACFLT funding for StrikePRO ended in 2001, prior to final verification, validation, & testing (VV&T) of the model. While there was never an independent review performed by competent, objective reviewers not employed by KAI, subject matter experts at the Naval Strike and Air Warfare Center (NSAWC) were able to review the methodology, with RADM Naughton asking for an in-depth series of briefings of the model's implicit modeling details, after which he committed NSAWC support to the project.

Methodology[1]

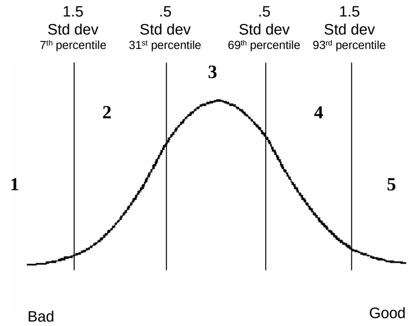
The general approach taken by the PRO models is to first identify resources that might plausibly affect operational proficiency, incorporating SME feedback on potential model inputs. Then, using as a benchmark some operational performance[2] that has been objectively captured, the analyst examines the correlation between higher resource levels and better performance.[3] PRO grouped resources into five categories:
People
Experience
Knowledge
Aptitude
Materiel



Within each of these areas, there might be several different resources, referred to as "bricks" and which could be combined into one overall unitless index for that resource area. This approach was designed to reduce the complexity of the model's multiple regression equations[4] and avoid multicollinearity[5] without introducing specification error.[6]

The method for combining bricks with different units was to assign them a value corresponding to which range of a normal distribution best represented the resource level. This binning approach is likened to computing an overall grade point average for a student taking diverse classes. Besides dealing with the challenge of different units for different independent variables, binning dealt with the issues associated with scale and lessened the need to use the natural logs of some variables to dampen variation in relation to the other variables. [7]

Bins



Raw Correlation

RS-6

Brick	Description	ATK
P1	# Pilots and NFOs onboard	-0.70
P2	Months of operational flying	-0.77
P3	Weighted CAT I/II/III Average	0.50
E1	# of FMC Flights last 30 days	0.56
E2	Avg. AAW/STW T-rating	0.98
E3	Avg. # of Flt Hrs per crew	-0.27
K1	# of Flt Hrs last 180 days	0.47
K2	# of Mission Flt Hrs last 180 days	0.63
K3	# of MC Flt Hrs last 180 days	0.67
M1	# of MC flights flown last 180 days	0.65
M2	# of DMMH/MC Flt Hrs last 180 days	-0.51
M3	Sum of ARI/Math/Mech/Gen Sci scores	0.86

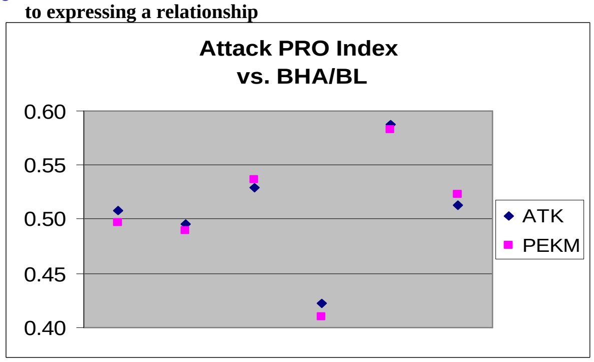
Each brick was examined for correlation against the operational performance data to allow estimation of which variables were most likely to exhibit a cause-and-effect relationship and estimate its individual strength.[8] With each resource area then represented by a single indexed value, a five-variable regression equation was sought that would approximate the observed operational performance values by balancing the impacts of the various independent variables. How well the regression equation, applied to the resource levels, could estimate the operational performance was then captured statistically.

In 2000, after analyzing the bomb hit assessment data on hand from six air wing rotations at Fallon, the StrikePRO model was using a four-variable equation[9] that explained 94% of the variations in BHA accuracy between air wings. For six data points, and a four-variable equation, the probability of that degree of "goodness of fit" occurring by chance was less than 5%, which is the commonly accepted threshold for statistical significance. While those results were promising, we realized that a population of six data points was not yet robust enough to ensure normal

Attack Model

> 95%

AT-6 For the first six events, a four-variable equation came very close to expressing a relationship



distribution and that time would be required to amass more data. Since there are only four to five Fallon detachments per year, we expected that it would take up to six more years before we either had a reliable model that could pass VV&T in an independent review or we had proven that the methodology was flawed and it was not possible to use cause-and-effect relationships to be able to estimate air wing bombing accuracy based on key resource levels.[10] COMPACFLT N-8 cancelled funding for the project when results were not forthcoming earlier.[11]

Assumptions

<u>Adjusted BHA/BF is the right performance metric</u>. As the dependent variable that StrikePRO attempts to estimate, based on NSAWC input, we use the number of bombs assessed as hits divided by the number of bombs an air wing plans to drop during its Fallon detachment. We adjust for no-counts for only two reasons: weather or range safety.

<u>Subject matter experts can recognize correlations that are cause-and-effect</u>. The point of developing a multiple-variable regression equation to estimate performance is to identify the resources that have a cause-and-effect relationship on proficiency and quantify the strength of that relationship – in short, to forecast estimates. With a stable equation that explained a large portion of variation and met statistical tests for significance, we would understand how much impact any one variable had on proficiency.

<u>Random effects will always introduce unexplained errors</u>. Even if all of the correct cause-and-effect relationships are identified and incorporated into the model's internal logic, there will be effects from random elements that will create a difference between estimated performance and observed performance. It is the responsibility of the analyst to review this error of estimation and determine whether it is within acceptable limits. There are statistical tests with widely accepted standards that suit this purpose.

<u>Normal distribution</u>. It is important to keep in mind that the PRO methodology depends on being able to deal with variation within each independent variable as if there were normal distribution. With only six data points, no statistician would yet be confident in predicting normal distribution. A good rule of thumb is provided by the Central Limit Theorem, that tells us getting a sample size of at least 30 air wings would be adequate.[12]

<u>Proficient units will demonstrate that proficiency, given sufficient time</u>. We know that experienced attack pilots can fling a bomb wide of the target on occasion and inveterate bombers can get a lucky hit. Over time, we expect to see observed performance in character with expected proficiency. With each air wing dropping somewhere between 50 and 300 bombs during a Fallon detachment, we trust that an air wing's BHA accuracy would not fluctuate much if it were to drop more ordnance (except that we would expect to see some improvement due to increased recent experience.) However, each air wing is made up of on the order of 50 pilots who drop the ordnance, so the sample size is deceptively smaller than it appears. Not every pilot will perform at his individual mean, but we expect the performances to offset for each air wing.

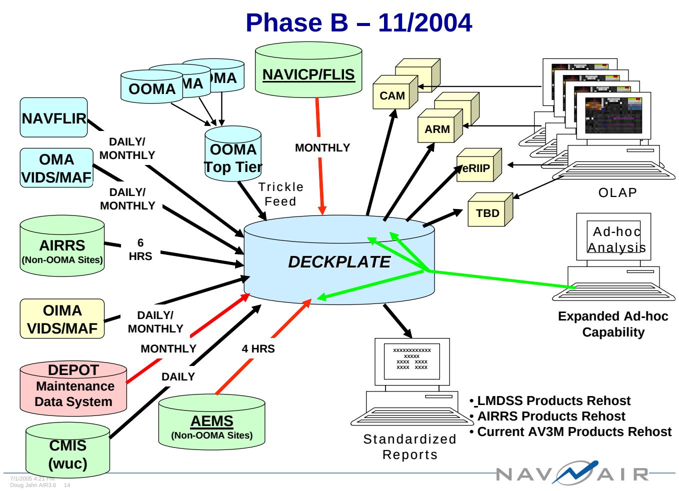
Challenges

The two most serious errors for regression models are:

- The error of omitted variables. Liaison with SMEs is required to ensure all possible cause-and-effect relationships are identified. Data from each of those independent variables must be sifted and tested for correlation before any variable is discounted. There are statistical tests that will identify when one or more significant variables have been left out.[13]
- An error in variable. Spurious data in either an independent variable or the dependent variable is a problem. When data containing biases or measurement error are used in a regression model, the accuracy of statistical explanation and prediction are affected. If the dependent variable (BHA/BF) contains error, but the independent variables are generally free of it, and the error is random, then the error will increase the errors of prediction while the estimated coefficients remain unbiased, which will mean the results have more variation. With a large sample size, the random errors will cause the average of all estimated coefficients to approach the actual coefficient. This is a lesser problem than the one caused by error in the independent variable. If the independent variables contain measurement biases, bias spreads through the error term in the estimate equation, which cannot be easily corrected. In its effect, this bias in the independent variable will cause the independent variable to be correlated with the error term.[14]

There were four sources for resource data. We experienced varying degrees of success in capturing data from each. Since the 24 air wing rotations at Fallon occurred between early 1996 and late 2001, we anticipated some difficulty in finding archived data that was up to nine years old.

- Aviation Data Warehouse. Innova maintains ADW for CNAF on its SIPRNET site. While we were unable to mine the data we needed from their website, they were able to send us archived data that was kept offline that we used to explore F/A-18 squadron SORTS reports just prior to their Fallon detachments. 19 of the requested 69 data points were missing, requiring us to make assumptions that call into question the validity of E2 data. We were unable to extract E3 data from this data.
- TRMS. Innova maintains TRMS for CNAF & CNAL on their SIPRNET websites. While we were unable to mine the data we needed from their website, they were able to send us archived data that was kept offline that we used to explore F/A-18 squadron personnel and manning status prior to their Fallon detachments. We were unable to reconstruct an adequately reliable picture of the personnel variables from the F/A-18 squadrons and consequently used none of the bricks within the P resource area.
- CACI. CACI provided flight hour data on the F/A-18 squadrons. The flight hour data from CACI differed from the flight hour data from ADW which differed from the flight hour data from the Naval Air Logistics Data Analysis (NALDA) program at NAVAIRSYSCOM. We estimate that there was an average error rate of 4% before we made assumptions that call into question the validity of K1 & M1 data.



- NALDA. By the end of the project, we were able to use the Decision Knowledge Programming for Logistics Analysis and Technical Evaluation (DECKPLATE) application to access the flight hour, Subsystem Capability Impact Reporting (SCIR), and maintenance databases -- Aircraft Inventory Readiness and Reporting System (AIRRS), Naval Aviation Maintenance and Material Management Yellow Sheet Reports (AV3M), Logistics Management Decision Support System (LMDSS), and Configuration Management Information System Work Unit Codes (CMIS WUC) at NALDA. Additionally, they were able to send CDR Ted Kaehler (CNAF N42) archived data that was intended to meet our needs. Unfortunately, the problem was not an inability to call up monthly data for specific Navy and Marine Corps F/A-18 squadrons from nine years ago. The problem was that about 4% of those months had no data reported or had data that was very obviously wrong. We used what judgment we could to massage the data, but recognize this calls into question the validity of E1 and K1 data.

Clearly, mining reliable resource data is both crucial to the StrikePRO methodology and difficult to do. NALDA's project officer made clear that NAVAIR recognizes data quality is an issue. They are working with TSD Orlando on an SBIR to address that issue. They documented several kinds of errors that they were aware existed in their databases that taken together called into question the validity of modeling using their data.

- Operational status codes conflicted, with ship fight hours logged while not deployed and deployed status but no ship flight hours logged in some cases.
- Poor documentation is affecting maintenance man-hours per flight hour calculations.
- Aircraft that were out of reporting status were being reported with flight hours against them.
- Aircraft listed at the wrong location.
- Incorrect "action taken code" used on 10% of cannibalization reports.
- Months where total flight hours were exceeded by the subset of total flight hours which represents strike training or air combat training.
- Months with no reported data at all.

As we groomed databases, we made assumptions that can be challenged and that call into question any relationships based on them if they are in error. In some cases, we wrote off a brick as entirely unreliable, rather than either a) assuming a constant for all air wings or b) possible inducing relative errors by wrong judgments. With this level of difficulty

cleansing the databases so that they could be used, it is unclear whether more time to work on the databases would have made significant improvements in the final results. Nonetheless, the CACI/L3 GSI team moved too slowly in the early months of the project, failing to anticipate how difficult and drawn out it would be to collect data once the data sources were identified.

Preliminary Findings

Here is a brick by brick review of the bricks used when we had six data points in our population and how they looked during this analysis with 24 data points in the population.

Raw Correlation

RS-24

Brick	Description	ATK
P1	# Pilots and NFOs onboard	
P2	Months of operational flying	
P3	Weighted CAT I/II/III Average	
E1	# of FMC Flights last 30 days	0.26
E2	Avg. AAW/STW T-rating	-0.49
E3	Avg. # of Flt Hrs per crew	
K1	# of Flt Hrs last 180 days	-0.09
K2	# of Mission Flt Hrs last 180 days	-0.26
K3	# of MC Flt Hrs last 180 days	-0.12
M1	# of MC flights flown last 180 days	-0.17
M2	# of DMMH/FMC Flt Hrs last 180 days	-0.29
M3	Sum of ARI/Math/Mech/Gen Sci scores	

- P1 – Number of pilots on board. Not modeled in 1999. Data not considered reliable enough to use in 2005.
- P2 – Number of months of operational flying per pilot. Not modeled in 1999. Data not considered reliable enough to use in 2005.
- P3 – Index expressing what portion of pilots in squadron were Cat I, Cat II, or Cat III. Data not considered reliable enough to use in 2005.
- E1 – FMC rate last 30 days times number of hours flown last 30 days. FMC rates cobbled together from multiple sources and considered reliable. Hours flown last 30 days had 4 of 69 months with data holes where assumptions could be challenged. This was the ONLY brick with correlation commending it to incorporation into a cause-and-effect model. At that, 26% correlation is not particularly strong. It suggests that this would be a minor variable alongside other more dominant variables.
- E2 – AAW/STW T-rating. With 19 of 69 months having data holes, suggests error in variable. Negative correlation for this brick is puzzling. It suggests that the better the T-rating of a squadron in STW & AAW before it goes to Fallon, the worse it does at bombing. That is counterintuitive.
- E3 – Flt Hrs/crew last 30 days. Data not considered reliable enough to use in 2005.
- K1 – Flt Hrs last 180 days. Was a candidate to be screened for multicollinearity with E3, K2, K3, & M1. Negative correlation for this brick suggests that the number of flight hours a squadron flew in the last 180 days had almost no impact on their proficiency at bombing.
- K2 – AAW or STW Mission Flt Hrs last 180 days. 8 of 138 months with data holes where assumptions could be challenged. Negative correlation for this brick is puzzling. It suggests that the more hours a squadron spent training in air-to-ground or air-to-air warfare in the 180 days prior to Fallon, the worse it did at bombing. That is counterintuitive.

- K3 – Mission-capable Flt Hrs last 180 days. Was a candidate to be screened for multicollinearity with E3, K1, K2, & M1. Negative correlation for this brick suggests that the number of flight hours a squadron flew in the last 180 days, times its mission capable rate, had a slightly adverse impact on their proficiency at bombing.
- A – No Aptitude bricks were modeled in 1999. Pilot AQT/FAR grades were considered unmineable. Pilots who had been varsity athletes in college was considered unmineable. The M3 brick was considered a candidate for reclassification as an aptitude measurement, rather than a materiel measurement, but the question was moot since GCT/ARI scores of maintenance men were considered unreliable in 1999.
- M1 – Mission-capable Flts last 180 days. Was a candidate to be screened for multicollinearity with E3, K1, K2, & particularly K3. Negative correlation for this brick suggests that the number of flights a squadron flew in the last 180 days, times its mission capable rate, had a slightly adverse impact on their proficiency at bombing.
- M2 – Direct maintenance Man-hours per flight hour last 180 days. All 69 months had data in them, but NAVAIR had singled out this calculation as particularly unreliable, in their judgment. (We know there are at least 17 of 69 data holes in flight hours per month.) That calls into question whether error in variable has made the data base unreliable. The negative correlation here indicates that, to a limited extent, the more maintenance that was required per flight hour in the last 180 days, the more proficient the air wing at bombing.
- M3 – The M3 brick was considered a candidate for reclassification as an aptitude measurement, rather than a materiel measurement, but the question was moot since GCT/ARI scores of maintenance men were considered unreliable in 1999 and again in 2005.

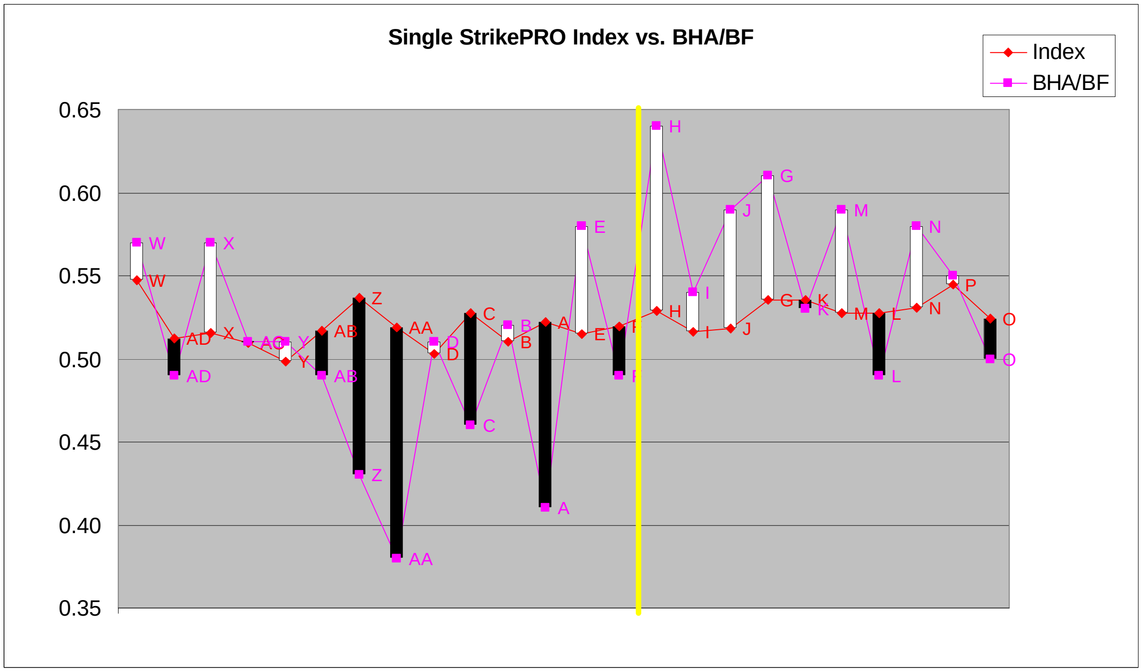
Insights

With positive correlation with only one brick, it would be nonsensical to attempt to derive a multiple regression equation to estimate BHA/BF. However, running the regression function in Microsoft Excel provides some analytic insight into the relationships.

First, using the developed regression equation, we can explain less than 4 per cent of the variation between the 24 air wing BHA/BF scores. For statistical significance at the 95% level, we would have needed to account for about 57% of the variation.

In a classic sign that one or more significant independent variables were left out, the impacts were captured by the intercept term. The low value of R square reinforces that conclusion. The Adjusted R square term suggests that we also included irrelevant variables, which can be spotted by their low t values. The E, K, and M variables are all shown to be irrelevant.

While the analysis allows us to be certain that we are missing some key variables, what we cannot tell is whether simply being able to mine all the bricks used in 1999 would provide a model that is able to adequately address variation in BHA/BF scores, or whether we would then still be required to perform sequential searches for unidentified factors that impact bombing proficiency.



Even without a reliable model, a look at the plotted data showed that something was fundamentally different about the relationships before and after mid-1999 (shown as to the left and to the right of the yellow vertical line). Prior to mid-1999, air wings generally performed poorly compared to estimates based upon their resource levels. After mid-1999, they generally outperformed expectations. We reexamined the data, attempting to run two separate equations – one for the first 14 points and a second for the last 10 points.

With an equation for just the first 14 air wings, we can account for 15% of the variation in BHA/BF scores. While this is certainly larger than the amount when we used a single equation for all 24 air wings, it is nowhere near the 73% we would need to be able to account for in order to indicate statistical significance at the 95% level. The low R square value and dominance of the intercept term indicate we have still left out one or more significant independent variables. The much lower adjusted R square and statistical insignificance of the E, K, & M terms, with accompanying low *t* values, indicate we have still included irrelevant variables.

Similarly, with a second equation for just the last 10 air wings, we can account for 13% of the variation in their BHA/BF scores. While this is certainly larger than the amount when we used a single equation for all 24 air wings, it is nowhere near the 82% we would need to be able to account for in order to indicate statistical significance at the 95% level. The low R square value and dominance of the intercept term indicate we have still left

out one or more significant independent variables. The much lower adjusted R square and statistical insignificance of the E, K, & M terms, with accompanying low *t* values, indicate we have still included irrelevant variables.

The original purpose of running the StrikePRO model for these 24 air wing rotations to Fallon was to provide a benchmark for the System Dynamics Readiness Model. However, if StrikePRO were undergoing VV&T, we would say this:

- Verification – This copy of StrikePRO failed verification. We were not able to build the model right, since we had unmet needs for resource data.
- Validation – This copy of StrikePRO failed validation. It does not adequately explain the variation in air wing scores in BHA/BF. It is unknown whether the model would have passed validation had it passed verification.
- Testing – This copy of StrikePRO failed testing. We know that the model has at least two major errors:
    - Error in variable. Spurious resource data could be causing compound errors.
    - Error of omitted variables. It is unknown whether the model would still have error due to omitted variables if it were able to pass verification and we could meet the resource data needs.

The bottom-line for StrikePRO remains ambiguous. We cannot reject the null hypothesis based on these results. To do so would be a serious Type I error if it turns out to be true – meaning that there are no cause-and-effect factors that affect bombing proficiency on Fallon detachments.

On the other hand, it could be a Type II error to accept the null hypothesis based on these results if the hypothesis is really false – meaning that there are really cause-and-effect relationships that impact air wing bombing proficiency on Fallon detachments, but we were just unable to disprove the null hypothesis.

The Road Ahead

So we cannot yet reject the null hypothesis, but it would be folly to accept it at this stage, based on the performance of a model that fails verification and contains known errors. One approach would be to gather operational performance data for more recent NSAWC detachments (or for a large number of SFARP detachments, or for a large number of COMPTUEXes, or for a large number of JTFEXes). The problem common to all of these sources for operational data is that only about four or five air wings go through training in any one year. No matter to which source we turn for robust data on BHA/BF, we probably need to collect about 30 or more air wing's worth of data, which may mean we need the corresponding resource data for about 6 to 7 years. It may be no small task to groom the resource data to the point where it can support modeling.[15]

One approach is using operational performance data drawn from all four sources, blending their BHA/BF so as to collect the data many times faster, this is a high-risk approach. But, if the character of the data is significantly different at any of these

locations, when we test for heteroskedasticity we'll likely be forced to segregate that data into a sub-population anyway. That would once again slow the rate of data accrual. It is possible that the advanced phase of SFARP and air wing Fallon detachment data may skew together into one sub-population and early SFARP, COMPTUEX, and JTFEX data may skew together into a second sub-population, based on two differing levels of fidelity in collecting target area information and defensive maneuvering prior to weapons launch. Additional consideration should be given to the fact that COMPTUEX and JTFEX collect nowhere near the same amount of BHA data, and their data would run a significant risk of not showing normal distribution.

Finally, there is the additional challenge that NSAWC is currently very protective of their air wing performance data and has not yet cooperated in providing recent operational performance data.[16] However, if cooperation can be gained from both NSAWC and the Strike Fighter Weapons School, and NALDA is able to report some success in grooming Navywide F/A-18 maintenance and SCIR databases, it may be possible to find corresponding performance and resource data from the most recent three years or so that do not introduce error. If that is possible, it may allow successful model verification. After that, validation could be attempted.

Endnotes and References

[1] For discussions that provide a non-statistician with a primer on multiple regression equation models, see Decisions by the Numbers: An Introduction to Quantitative Techniques for Public Policy Analysis and Management, by Dipak K. Gupta (Prentice Hall: Englewood Cliffs, NJ), 1994, Chapters seven and eight.

[2] Initially, StrikePRO used the number of bombs assessed as successfully hitting the target divided by the number of bombs that an air wing launched while on its Fallon detachment – BHA/BL. Later, on the recommendation of NSAWC SMEs, we changed to BHA/BF – or the number of bomb hits divided by the number of bombs the air wing planned to drop (or bombs fragged).

[3] Actually, PRO starts by having SMEs craft an Operational Sequence Model for strike warfare, and then propose possible causal factors that might have an impact on the observable outcome of interest – in this case BHA/BF. Analysis begins with a null hypothesis, stating that there is no cause-and-effect relationship between two variables. If further analysis can then disprove the null hypothesis with at least 95% certainty, we conclude that one of them is dependent on the other.

[4] This would limit the PRO model to a regression equation to estimate least-squares variation with five variables, six coefficients, and five degrees of freedom. The regression tool in Microsoft Excel is easily capable of this level of complexity.

[5] Multicollinearity occurs when two variables, thought to be independent, turn out to be correlated. The estimated coefficients of those two variables become extremely sensitive and will change dramatically each time the equation is run with a slightly different set of independent variables, even though the measure of goodness of fit will not be affected. When the two variables do not have opposite expected signs, and are not significantly different in magnitude, combining them into a newly created variable that contains information from both variables will eliminate the multicollinearity and stabilize the equation. Using binned values in PRO eliminates the risk of the two variables differing in magnitude. Running a simple correlation test prior to deriving the regression equation will alert the analyst to opposite expected signs.

[6] Specification error is incorrectly identifying the significant independent variables in the model. Omitting a significant variable is a serious problem for a model that is intended to forecast estimates based on cause-and-effect relationships. Fortunately, there are statistical tests that can identify an omitted variable, although the only correction for this error is a deeper understanding of the nature of the dependent variable that suggests another possible independent variable. Adding an irrelevant independent variable is a minor problem for the model. Statistical tests will identify irrelevant variables and the correction is as simple as eliminating them from the model.

[7] Even after using bins to deal with scaling factors, transforming a variable by taking its natural log may still be necessary if variation is known to be non-linear. The rule of thumb is to seek the Best Linear Unbiased Estimator (BLUE), unless theory, common sense, or SME experience suggest otherwise. Therefore, PRO methodology is to avoid using natural log transformations solely because they produce a better fit and lacking sound rationale for why the specific relationship is non-linear.

[8] Strong positive correlation, or numbers approaching +1.0, means that as the independent variable increases, so does the dependent variable. Strong negative correlation, or numbers approaching -1.0, means that as the independent variable increases the dependent variable decreases. Weak correlations, or those approaching 0, mean that whether the independent variable increases or decreases has little effect on whether the dependent variable increases.

[9] We had been unsuccessful in isolating an Aptitude variable with significant impact on BHA/BL.

[10] We considered augmenting the number of opportunities for recording operational performance data by also recording BHA/BL at Strike Fighter Advanced Readiness Program (SFARP) detachments, COMPTUEX, or JTFEX, which would have more quickly built a population of at least 30 observations. However, BHA data is greatly influenced by the ability of the aircrew to fight his way to the target past surface-to-air threats and air-to-air threats. The fidelity of training in that high threat environment is so much higher during NSAWC detachments that we could not rule out heteroskedasticity if we observed at the other events. In other words, we would be seeking to combine apples and oranges that good statistical review would later cause us to track separately anyway. Our sub-populations would reach the target number of 30 observations no faster than would a single population of purely NSAWC data – which was by far the most reliable.

[11] Verification is the process of determining that a model or simulation implementation accurately represents the developer's conceptual description and specifications. Model verification is substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy. Model verification deals with building the model right. The accuracy of transforming a problem formulation into a model specification or the accuracy of converting a model representation from a micro flowchart form into an executable computer program is evaluated in model verification. Validation is the process of determining the degree to which a model or simulation is an accurate representation of the real world from the perspective of the intended uses of the model or simulation. Model validation is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the M&S objectives. Model validation deals with building the right model. An activity of accuracy assessment can be labeled as verification or validation based on an answer to the following question: In assessing the accuracy, is the model behavior compared with respect to the corresponding system behavior through mental or computer execution? If the answer is "yes" then model validation is conducted; otherwise, it implies that the transformational accuracy is judged implying model verification. Testing is ascertaining whether inaccuracies or errors exist in the model. In model testing, the model is subjected to test data or test cases to determine if it functions properly. "Test failed" implies the failure of the model, not the test. A test is devised and testing is conducted to perform either validation or verification or both. Some tests are devised to evaluate the behavioral accuracy (i.e., validity) of the model, and some tests are intended to judge the accuracy of model transformation from one form into another (verification).

[12] For a significance level of at least 95%, meaning that there is less than 5% likelihood that correlations are due to chance, the Fisher ratio test enables us to determine that a model with five variables would have to account for at least 49% of the variation between the estimated and observed performances of 30 air wings.

[13] If one or more significant variables have been left out, statistical analysis will show a very low value of r-square and the intercept term coefficient will capture most of the impact in the multiple-regression equation. The only correction is to improve understanding of the dependent variable. Sequential searches, refining the model over time through liaison with SMEs, are good protocol. Stepwise regression or "fishing expeditions," where we select independent variables based solely on the strength of association and not on logical causality, are flawed techniques since they are built on co-occurrence that is unlikely to persist for the long term.

[14] See Gupta, pp 253-4.

[15] The NALDA project officer indicated that their problem with Fleet database errors continued at least up to data recorded in 2003. The results of the SBIR intended to groom that data so that it can be used for modeling may eventually tell a lot about what is really possible with the StrikePRO methodology.

[16] When COMPACFLT pulled the funding on StrikePRO in mid-2000, NSAWC had provided in-depth operational performance data on 16 air wings and BHA data on another 8 air wings, but had not gotten any feedback on how well resource data allowed KAI to forecast an estimate of operational performance. Subsequent to that, it has been challenging to get NSAWC participation in outside analysis of their performance data.